

Setting a bad example on AIDS

South Africa has a unique opportunity to play a key role in creating effective health-care systems in sub-Saharan Africa. It must not squander this opportunity by rejecting the help offered by science.

It is impossible to claim that South Africa has been turning its back on its AIDS problem. The disease is now so endemic, and absorbs so much of the health budget at both federal and provincial levels, that it has rapidly become one of the country's top health priorities. All the more reason, therefore, to be concerned at the government's recent decision not to supply the drug AZT to pregnant women infected with HIV (see *Nature* 396, 504; 1998). Whatever the appeal of the argument that the money for this drug could be better used on prevention, the move threatens to undermine South Africa's credibility as an effective platform for improving the health of the whole of sub-Saharan Africa.

The statistics are horrifying enough. Nationally, officials estimate that 1,500 individuals are infected every day in South Africa, and prenatal tests have revealed that about 16 per cent of all pregnant women are HIV positive. The situation has been developing rapidly; the result, inevitably, has been an equally rapid increase in the number of children born with the infection, with the drain on both medical and social resources that this inevitably creates.

Yet Nkosazana Zuma, the country's soft-spoken but tough-minded health minister, has steadfastly refused to use government funds to pay for AZT for infected mothers. Her reasoning has a superficial logic; even though a three-month course of AZT costs only about 1,500 rand (US\$250), the same money spent on prevention could — at least in theory — save many more than one (or two) lives. But this cold calculation fails to take into account a different sum: when the cost of treating a mother for this period is compared with that of providing medical treatment for an infected child — one estimate places this at more than 50,000 rand — the savings are undeniable.

Another agenda seems to be at work, one based on an implicit distrust of certain aspects of Western biomedical practice, particularly when corporate interests (as expressed through the involvement of major pharmaceutical companies) are involved. Zuma, a physician by

training, has already demonstrated her disdain for some aspects of this practice through her vigorous promotion — circumventing the country's Medicines Control Council — of a group of researchers at the University of Pretoria who claimed last year to have produced a local cure for AIDS (see *Nature* 386, 6; 1997).

Such distrust is not totally without foundation; pharmaceutical companies are, after all, driven by private profit. And Zuma, an influential member of the ruling African National Congress, has many supporters who back the principles on which she has been standing firm (a ban on tobacco advertising has been another, equally controversial topic). The cabinet, for example, has publicly endorsed her decision not to provide pregnant women with AZT, which would have had to have been bought from Western drug companies.

But there is also a downside to Zuma's rigid stance. Some argue that it could create a situation in which drug companies will be reluctant to invest in preventative strategies if these are to be given a low priority. Others point out more concretely that children who could have been saved are being condemned to a painful, untimely and unnecessary death. The situation has already enraged some Western researchers so much that they are threatening to boycott the next (13th) international AIDS conference, due to be held in Durban in 2000. This would be a mistake; boycotts by a few individuals seldom make an effective political weapon.

But the sentiment behind the threatened boycott is correct. Whatever Zuma's suspicions of Western drugs and drug companies, withholding affordable treatment from individuals who would benefit enormously from it (a study in Thailand showed that the chance of a baby being born with HIV almost halved when the mother was given AZT) verges on the immoral. It is time for the government to rethink its stance. Little in either practical or financial terms would be lost. Much in terms of scientific and humanitarian credibility would be gained. And many lives would be saved. □

Protest in Paris

The successful reform of French science needs a greater commitment to openness and consultation.

"No one can simply bring together a country that has 265 kinds of cheese"; General Charles de Gaulle's exasperation with the French is probably shared this week by Claude Allègre, the country's science minister. This follows a rebellion in the scientific community over plans to put the country's fundamental research agency — the Centre National de la Recherche Scientifique (CNRS) — on the road to becoming primarily a funding agency for university-based research, and giving universities joint responsibility for the running of CNRS laboratories (see page 607).

But Allègre has only himself to blame for the grass-roots challenge to his reforms. Over the past year, he has repeatedly — and unfairly — attacked CNRS as being solely responsible for all the woes of French science, and unnecessarily antagonized the national committee for scientific research, a sort of 'parliament' of scientists, by publicly describing it as a hotbed of nepotism and bureaucracy. He has also avoided direct consultation with representatives of the scientific com-

munity, pursuing reforms behind closed ministry doors.

It would be a mistake to dismiss the backlash that Allègre has provoked as merely a reflex defence of the status quo. It is not. The scientific community itself recognizes the need for change, and accepts that the proposed reforms contain some good ideas. But it is also worried that they smack of haste and authoritarian technocracy. In contrast, many researchers still cherish an idea of French science that involves collective input. These are making a legitimate demand for greater consultation in reaching agreement on what would constitute comprehensive and meaningful reforms.

Debate should not become an excuse for inaction, however, and Monday's show of force needs to translate quickly into concrete proposals for change. For his part, Allègre now needs urgently to find ways of involving the scientific community more broadly and openly in dialogue on how this can be achieved. If he is unwilling, his close friend, prime minister Lionel Jospin, should find a new science minister. □



More foreign research students stay in US to cash in on boom

[WASHINGTON] Foreign research students who receive their doctorates in the United States are more likely than ever before to stay, taking advantage of the booming US economy to land postdoctoral fellowships or plum jobs in industry, according to figures collated by the National Science Foundation (NSF).

A survey has revealed that the number of students from Europe, Asia, Canada and Mexico obtaining PhDs in the United States rose sharply from 2,400 in 1985 to more than 8,000 in 1996, the last year for which complete data were available. But the number of these who say that they "plan to stay" in the United States grew even more rapidly, rising from 50 per cent to 72 per cent of the total (see graph).

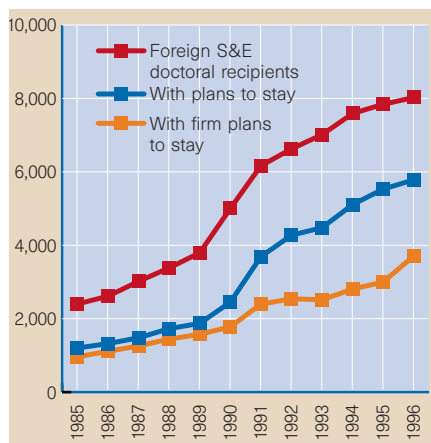
The number who say they have "firm plans to stay", meaning that they had received a firm job offer on receiving their degree, also reached a new high of 46 per cent in 1996. The NSF says that the response to this question is a good indicator of how many of the students will remain in the United States.

Students from China — by far the largest group, now accounting for more than one-third of the total — are the keenest and most likely to stay in the United States. Since Congress passed the Chinese Student Protection Act in 1992, granting most of them permanent residency status, the number saying that they hoped to stay has exceeded 90 per cent. By 1996, the number with firm plans to stay had reached 57 per cent.

Over the entire period surveyed, the students next most likely to stay were those from India, of whom 79 per cent plan to remain. Among students from European countries, those from the United Kingdom were by far the keenest to stay, with almost 70 per cent planning to do so. Of the other countries that send a large number of graduate students to the United States, about half of the students from Canada, Germany and Greece plan to remain, as do fewer than half of those from Taiwan and South Korea.

More than half of the British postgraduate students had received a firm job offer — more than any other nation except India — and 8 per cent had an offer to teach at a US university, almost twice the success rate of students from any other country in the study.

Students in the life sciences are the most likely of all to stay in the United States, with more than half reporting firm plans to stay, almost all of them planning to continue with postdoctoral study. Fewer of those studying the physical sciences, engineering and computer science plan to, and many of those who do will move directly into jobs in industry.



Staying Stateside: rising numbers of foreign science and engineering students hope to remain.

The study, which was published last month, excluded the smaller number of students who come to the United States from South America, Africa and the Middle East. The NSF says that the increase in research students choosing to remain in the United States is largely attributable to the sharp increase in the number of students from China and their tendency to stay after graduation.

"Although the increase in the 'stay rate' of Chinese students is largely responsible for

the jump in the overall rate of those planning to stay, students from other countries are also contributing to the growing number of those who plan to stay," says the report. "Undoubtedly the increase is a reflection of rising employment opportunities."

NSF officials believe that the underlying upward trend has continued since the last data were collected in 1996. "It looks like a pretty strong trend and I doubt that it will change," says Jennifer Bond, director of the NSF Science and Engineering Indicators Program.

Bond says that the strong US economy, combined with economic problems in Asia, will continue to encourage graduating research students to stay in the United States. But she points out that acute economic problems in some countries may mean that they send fewer students to the United States for postgraduate education.

In the mid-1990s, concern about unemployment among those with PhDs led to complaints from young US scientists about the number of foreign students being encouraged to study at US universities. But these complaints have been muted during the recent economic expansion, which has seen unemployment in the United States fall to the lowest rate for 30 years. **Colin Macilwain**

UN stands up for East German scientists

[MUNICH] The United Nations Committee on Economic, Social and Cultural Rights in Geneva last week noted "with alarm" the small number of scientists from the former German Democratic Republic (GDR) who have been re-employed after German unification.

The issue had been brought to the attention of the committee by human rights organizations and trade unions in Germany. The committee was unconvinced by the official German denial that discrimination against state officials and others associated with the former communist regime had ever taken place. The committee has called on the German government to provide compensation, employment or pensions "as an act of national reconciliation".

But German officials challenge the committee's conclusion that the scientists have been discriminated against politically. They argue that its statement claiming that most of those affected "may have been dismissed from their positions for political rather than for professional or economic reasons" is misleading.

Marion Bimmler, for example, head of the staff council at the Max Delbrück Centre for Molecular Medicine in Berlin, argues that "not more than 3 per cent" have been dismissed for political reasons, such as collaboration with the Stasi, the GDR's security service.

Nevertheless, only an estimated 15 per cent of the 24,000 scientists employed by the GDR's Academy of Sciences found permanent academic posts after the academy was dissolved in 1990. Each had to reapply for jobs in the new research institutes established along western lines, and many were unsuccessful in competition with candidates from west Germany and abroad (see *Nature* 370, 240; 1994).

Government-financed support programmes rescued some from immediate unemployment, but these finished in 1996.

Bimmler hopes that the government will provide more money to re-employ the scientists. She says that a new reintegration programme "would be much more helpful than sending competent people home with a pension". **Quirin Schiermeier**

Budget block on US bid to rejoin Unesco...

[WASHINGTON] A bid by the State Department to renew US membership of the United Nations Educational, Scientific and Cultural Organization (Unesco) has been blocked by the White House budget office.

The State Department had opened its effort to secure US re-entry with a request to the White House's Office of Management and Budget (OMB) that \$22 million — three months' contributions to Unesco — should be included in the administration's budget request to Congress for the fiscal year 2000.

But the OMB, which is said to remain sceptical of the chances of such a request getting past the Republican-dominated Congress, has omitted it from its advice to President Bill Clinton on the budget.

Various Democrat members of Congress, scientific bodies and organizations sympathetic to US re-entry into Unesco, and angered by the OMB's move, are appealing to Clinton to reinstate the Unesco allocation in his budget document, which will be submitted to Congress at the end of January.

Esteban Torres (Democrat, California), for example, a former US ambassador to Unesco, has written to Clinton advising him that the United States "can no longer afford to remain outside of Unesco".

Supporters of US re-entry argue that dropping the budget request could damage US interests in international science and education policy. This includes participation in the election of a new director general for the agency; plans for Unesco's post-2000 strategy; and preparations for and participation at next year's Unesco World Conference on Science in Budapest (see *Nature* 396, 299; 1998).

John Fobes, a former deputy director general of the agency who now chairs Americans for Universality of Unesco, says that the conference is too important for the United States to participate as an observer. "A mere announcement [that the United States is preparing to rejoin Unesco] would make a lot of waves," says Fobes. It would, he adds, effectively change the country's status from observer to full member.

The United States under Republican President Ronald Reagan walked out of Unesco in 1984 complaining that the agency was being mismanaged and lacked a clear focus.

Clinton gave a public undertaking on the fiftieth anniversary of Unesco in 1995 that the United States would rejoin, arguing that he was satisfied that key management

reforms had been implemented. But he gave no clear indication of timing.

The State Department's decision to push for US re-entry at this time has, however, surprised some close observers of relations between the United States and the United Nations, as Unesco is not considered to be top priority in the list of pending issues between the two.

Reversing the US withdrawal from the United Nations Industrial Development Organization (UNIDO) in 1996 is considered to be a higher political priority than rejoining Unesco.

Republican opposition has also stalled a proposed settlement in Congress of the United States' unpaid arrears to the UN, which are close to \$1 billion. One UN observer in Washington DC says that additional requests for UN contributions "will be received quite violently in some quarters".

The State Department has a strong ally in its efforts. Britain chose to rejoin last year after a similar absence, and is believed to be encouraging the United States to do the same. "Don't underestimate the power of [Tony] Blair going in," says one senior Unesco official. "The United States does not like to see itself alone."

Ehsan Masood

... as Australian minister hits out at agency over heritage sites

[SYDNEY] Australia's environment minister, Robert Hill, has come into conflict with non-government scientists over whether extra steps are needed to protect two of Australia's main environmental attractions.

The government is facing opposition to its decision to allow uranium mining in the Kakadu National Park in the Northern Territory. It is also under fire for its failure to take a firm stand on global warming, which, argue critics, is threatening the health of the Great Barrier Reef.

The government is rejecting demands that these areas, already featured as World Heritage Areas by Unesco, be placed on the Unesco 'endangered' list.

One of the disputes arose from a Unesco investigation into the damage that could be caused by the development of a uranium mine and mill at Jabiluka in Kakadu, which was approved by the government after years of opposition.

Environmental activists, the local aboriginal Mirrar people and many scientists, joined last week by the Australian Catholic Social Justice Council, argue that the mine would irrevocably damage the World Heritage values of the park.

The Unesco mission concluded that the park "is exposed to serious threats which are placing it under both ascertained and



Coral crisis: researchers have raised the alarm at Unesco over threats to the Great Barrier Reef.

potential danger". Hill, a lawyer, has accused the mission of bias, lack of balance and factual errors. The mission, led by Francesco Francioni, an Italian environmental lawyer who also chairs the World Heritage Committee, recommended that the Jabiluka project should not proceed.

In its report, presented to the Bureau of the World Heritage Committee at its meeting in Kyoto, Japan, two weeks ago, the mission said that its recommendations "should be implemented in a spirit of full transparency and public consultation in Australia". But the committee offered

Australia six months in which to present fresh evidence against a formal "threatened" label.

Construction of the uranium mine has been proceeding rapidly following a High Court rejection of an appeal by aboriginal groups.

In a separate issue, Hill characterized as "not compelling" recent evidence linking world-wide coral bleaching to rises in sea surface temperatures (see *Nature* 396, 10; 1998). His statement in parliament followed the release last month of a report on the status of the world's coral reefs by the Global Coral Reef Monitoring Network, an initiative of Al Gore, the US Vice-President.

The report concludes that reefs are "under considerable stress and are experiencing considerable damage" in association with rises in temperatures.

Frank Talbot of Sydney's Macquarie University, director emeritus of the Smithsonian Museum of Natural History in Washington DC, has joined other reef researchers in appealing to Unesco to re-examine its World Heritage listing of the Great Barrier Reef as a result of the effects of land clearing, coastal developments and overfishing. Otherwise, they predict "the slow death of the greatest exemplar of coral reefs in the world".

Peter Pockley

France's scientists rally to oppose Allègre's reforms

[PARIS] Plans by Claude Allègre, the French science minister, for a profound reform of the country's research system are hanging in the balance this week after a strong backlash from the scientific community.

The issue came to a head at an unprecedented meeting in Paris on Monday (14 December) of the 800-member National Committee for Scientific Research, the 'parliament' of the country's scientists which plays a major role in evaluating laboratories and administering recruitment.

The meeting attacked the reforms as "ill conceived" and overwhelmingly rejected the way in which Allègre, formerly the director of the Institut de Physique du Globe in Paris, has tried to impose them on the scientific community with minimum consultation.

The meeting was organized in large part as the launch of a national debate on science, felt by many scientists to be a necessary prerequisite for reforms, but repeatedly refused by Allègre. This move was backed by Catherine Bréchnac, director general of the Centre National de la Recherche Scientifique (CNRS), and senior CNRS officials. Hubert Curien, a much respected former Socialist minister of research, also endorsed the move as a "very good initiative" that "opened up the debate".

The plenary meeting of the committee — two-thirds of whom are elected by scientists — was the first in its 50-year history to be convened by scientists themselves. The massive turnout of more than 700 members was prompted by what many scientists fear are plans to dismantle Europe's largest fundamental research agency, the 27,000-strong CNRS, and to increase central control over research (see *Nature* 395, 729–730; 1998).

The atmosphere in the packed amphitheatre of la Maison de la chimie was one of rebellion, reminiscent of the student meetings of 1968. Usually reserved middle-aged academics and research administrators battled for microphones. Vincent Courtillot, Allègre's representative at the meeting, was repeatedly heckled, jeered and laughed at as he defended his minister's actions and plans.

The committee's challenge stems partly from its desire to have a say in the reforms. Although it sees itself as the representative of the scientific community, it has been excluded from discussions by Allègre. The meeting signalled a new assertiveness on the part of the committee, which had been criticized by Allègre as rife with nepotism and bureaucracy.

The minister has sought to create alternative evaluation structures within his ministry. But the thrust of the meeting was that a group of officials is no substitute for the col-



Curien: defended CNRS against Allègre's attack.

lective expertise of some 1,000 researchers, and that the reforms would fail unless they took into account the experience of scientists and a long-term view of research needs.

The resistance to Allègre's approach was illustrated by opposition from both CNRS and university researchers to his proposals to transfer much of the responsibility for the work of CNRS laboratories to the universities.

A succession of speakers accused the minister of risking the break-up of agencies such as CNRS that worked, while failing to address fundamental problems such as the weakness of university research.

Claude Cohen-Tannoudji, the 1997 Nobel prizewinner in physics, accused Allègre of naivety in trying to impose the US model of competitive research universities directly on France. He argued that the weaknesses of research in French universities that led to the creation of the CNRS in the 1930s were still alive and well, and that any reduction in the strength of the agency's laboratories would damage French science as a whole.

Curien also argued that French universities were in general no substitute for CNRS. "The universities are not ready yet; in 15 years perhaps, but for the moment, no."

Curien added that Allègre's attacks on the CNRS were unjustified, and that the major weaknesses of research — such as technology transfer and the mobility of researchers — required a comprehensive analysis of the entire research and industrial landscape.

Bréchnac was equally critical. "Our minister, with his familiar brusqueness, but with conviction, tells us we need to 'move', and he is right," she said. "But we need to know where we are supposed to go."

The mood of the meeting was that scientists are keen to see improvements in the research system, but want these to be well thought out and introduced gradually by the research agencies themselves. For the moment, though, the situation appears deadlocked. Flush with its success in mobilizing the scientific community, the national committee appears in no mood to modify its opposition.

But Courtillot reaffirmed the ministry's intention to proceed with reforms, warning that "organizations that refuse to adapt are condemned to disappear".

Declan Butler

UK set to back industrial spin-offs from research

[LONDON] Proposals to encourage the commercial exploitation of scientific research in Britain were expected to be given high priority in a government white paper (policy document) on economic competitiveness, due to be published this week.

The white paper was expected to propose changes in the way in which the Department of Trade and Industry (DTI) spends its annual budget of £1 billion (US\$1.7 billion). It was due to be unveiled by Peter Mandelson, Secretary of State for Trade and Industry, the cabinet minister responsible for science.

The measures are designed to help narrow the gap between the United Kingdom and its industrial competitors in terms of ability to benefit commercially from science and technology. Despite its position among the top few countries in terms of scientific output, the United Kingdom came thirteenth of 17 countries ranked in terms of their effectiveness in exploiting science in a survey by the Harvard Business School that was released last week.

Since taking up his post in a cabinet reshuffle earlier this year, Mandelson has been keen to find ways of encouraging partnerships between industry and the academic community, turning universities "from ivory towers into business partners".

For example, he is known to be enthusiastic about the Teaching Company Scheme run by the Engineering and Physical Sciences Research Council, which provides financial support to graduates working in companies on projects supervised jointly by the companies and universities. The scheme is expected to receive extra backing in the white paper.

Faraday Partnerships, which aim to improve the interaction between the science, engineering and technology base and industry, may also gain government support for their expansion plans.

Industry was hoping for changes to the tax system to make it easier for high-tech start-up companies to raise venture capital. And Mandelson has said he is keen to promote clusters of science-based businesses near universities.

Some of these proposals are likely to be financed by a reorganization of DTI funding of business initiatives. Rather than providing more broad-based regional aid, the department is keen to target funds on clusters of high-tech, 'knowledge based' industries.

The white paper follows a statement last month by the Chancellor of the Exchequer, Gordon Brown, that the government is considering offering a tax credit for small and medium sized companies yet to make a profit (see *Nature* 396, 100; 1998).

Natasha Loder

Claims of hidden agenda on French waste

[PARIS] The French government has announced plans to create a deep storage centre and two underground laboratories for studying the disposal of nuclear waste in clay and granite localities.

But critics say the plans reveal the government's long-term intentions on the disposal of nuclear waste, and as such conflict with a resolution to keep such decisions open.

The plans, revealed last week, are among a collection of measures relating to the French nuclear industry, including an attempt to improve the transparency of the industry and the organization of its security.

The laboratory to investigate clay sites will be set up at Bure in Meuse, eastern France. According to the National Agency for the Management of Radioactive Waste (ANDRA), it will begin operation in 2003.

The location of the second laboratory, on a granite site, has not yet been chosen. La Chapelle-Bâton in Vienne, western France, near Poitiers, has been suggested by ANDRA, but is no longer being considered.

Preliminary studies showed that the geological characteristics of the site will probably prevent it being used for waste disposal.

The underground repository, which will be less than 20 metres deep, is likely to be set up in Gard in the south of France, after the site has been "scientifically verified".

A report from the Commission on Atomic Energy has indicated that this type of storage is possible. But it "should be accompanied by a programme of research, in particular on methods of cooling".

The Green party, who are members of France's coalition government, sees the decision to drop La Chapelle-Bâton as an indication that the two laboratories will be not only the location of preliminary studies but also the site of future waste-disposal facilities.

The Greens have accused the government of giving in to the nuclear lobby on this matter. But the government insists that it plans to stick to the 'loi Bataille' of 1991, and the concept of 'reversibility', maintaining the chance of keeping access to the stored waste.

The 1991 law specifies that all options for storing nuclear waste should be investigated to ensure that parliament is in full possession of the facts about alternatives when it chooses the method of storage in 2006 (see *Nature* 390, 322; 1997).

A report of the National Assessment Committee, an independent body, has established a clear link between the type of waste and the type of storage. In particular, the report specifies that "deep storage, involving the triple barrier of a container, a constructed barrier, and then geological barriers, makes reversibility very difficult, if not improbable in the long term". It also argues that "reversibility is easy in subsurface repositories".

Despite this, ecologists are concerned that both the committee and the government are speaking about "definite storage" for the so-called 'B wastes' with medium radioactivity and long lifetimes, without specifying precisely what this means, but contrasting it to subsurface storage.

Eric Glover

Search hopes to detect laser signals from extraterrestrial life

[BOSTON] Researchers at Harvard University and the University of California at Berkeley have begun combing the heavens for flashes of laser light emitted by alien civilizations.

In October, the Harvard group, led by physicist Paul Horowitz, mounted a camera on a 61-inch optical telescope at the Oak Ridge Observatory in Harvard. The device rides 'piggyback' on a survey of 2,500 Sun-like stars, part of a search for extrasolar planets run by the Harvard-Smithsonian Center for Astrophysics, and uses about a third of the light from the telescope in its search for short laser pulses. They have found 30 intriguing signals from the nearly 1,000 stars scanned so far, but all were written off when they failed to repeat.

The Berkeley effort, led by astronomer Dan Werthimer, takes a similar approach. Early next year, they will connect a laser detector to a 30-inch automated telescope at Berkeley's Leuschner Observatory that will observe 2,500 nearby stars. Meanwhile, Berkeley astronomer Geoff Marcy, co-discoverer of many planets outside the Solar System, is looking for laser signals among the data from his planet search.

Previous searches have almost all used radioastronomy. But according to Horowitz, recent technical advances have made optical SETI (search for extraterrestrial intelligence) a viable approach that is rapidly catching up with radio SETI. Whereas radio waves can better penetrate an atmosphere, lasers carry information more efficiently, and are not dispersed by the



interstellar media to the same extent as radio waves.

Interference — a growing problem for radioastronomers on Earth because of the proliferation of cellular phones and communications satellites — is not a problem in optical wavelengths. With laser power output doubling every two years, Horowitz says, "we can now send out a pulse across the Galaxy more than a 1,000 times brighter than the Sun".

He believes a 'pulsed optical beacon' from a more advanced civilization could outshine its parent star by a factor of a million. These fast-talking extraterrestrials could signal to 1,000 stars in about a minute.

One of the biggest challenges currently facing SETI enthusiasts is deciding where in the heavens — and in the electromagnetic spectrum — to look. "If I were transmitting

laser pulses, I'd do it in near-infrared rather than optical," Horowitz says, admitting that it's hard to know how to optimize the search. "Maybe they're sending 'zeta waves' and we haven't discovered them yet."

Even after searching for 20 years with no significant results, Horowitz is still optimistic. "We're the first generation that could realistically establish contact with another civilization, which would be the greatest event of all time, and almost no one is trying," he says.

"Earthlings are just learning how to do this," adds Werthimer. "The Universe could be filled with radio or laser signals shooting around, and we wouldn't have a clue." Searching for alien life is like looking for a needle in a haystack, he adds, but right now "Earthlings are just exploring the corners of the haystack. Eventually we will be able to search a large chunk of the haystack."

The new optical searches were stimulated partly by a workshop held last year by the California-based SETI Institute to design searches for the next two decades. "Most of what they are planning to do is more radio searches that cover larger chunks of the spectrum," says Horowitz. "That makes sense, but it doesn't mean we can't try something else too."

Horowitz was also inspired by a talk given in 1995 by Charles Townes, who won the Nobel prize for his work on lasers and masers, promoting the idea of optical SETI. "I was blown away thinking: 'How did I miss this?'" he says.

Steve Nadis

European physicists split on new magnet

[MUNICH] Physicists are divided over whether Europe's next large magnet should provide a continuous field, which would offer a wide variety of uses, or a pulsed field, which would give a much more powerful field but for a relatively short duration. Or, indeed, whether new high-field magnets are needed at all.

A report from the European Science Foundation (ESF) on the scientific cases for developing both types of magnet argues that the most urgent need is for a long-pulse (10–100 millisecond duration) 100-tesla magnet facility, as US scientists are taking the lead in the development of pulsed fields.

But not all scientists are convinced that developing stronger magnets is a priority for Europe. Scientists in Germany and Britain are being polled about their opinions.

The ESF report acknowledges that the highest demand is for continuous rather than pulsed fields, but says that such facilities in Europe are already state of the art. The case for developing continuous magnetic fields should be reconsidered in a few years, it says.

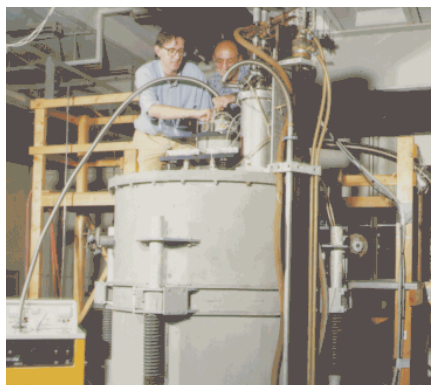
But Europe is almost certain to fall behind the United States in the provision of high-field, long-pulse magnets, says the report. The world's first 100-tesla long-pulse magnet is expected to be operational at the National High Magnetic Field Laboratory in Los Alamos, New Mexico, next year.

A two-year technical study for a similar European magnet, funded by the European Commission, has just been completed by experts brought together by the ESF. They recommend building a prototype by 2003, at an estimated cost of between ECU50 million (US\$60 million) and ECU100 million.

Research into semiconductors, low-dimensional structures, magnetism and metals, superconductors and some areas of chemistry, including optical spectrometry, would benefit from such a facility, says the report. A 100-tesla field would raise the energy of a free electron to the level existing between interacting electrons, allowing complex quantum effects to be studied.

But some European physicists believe there are too few potential users of powerful magnets to justify such costs. About 500 scientists use Europe's largest continuous high-magnetic-field facilities in Grenoble, France, and Nijmegen, The Netherlands, which are being upgraded to generate magnetic fields of 40–45 tesla. In contrast, between 200 and 300 use pulsed-field sources.

Proponents of investment in stronger magnets say there would be more users if new facilities were available. "If higher-field magnets are built, interesting and unanticipated science will undoubtedly flow from them," says Mike Springford, professor of physics at the University of Bristol, United Kingdom, who chaired the scientific committee.



Big attraction: the 25-tesla Nijmegen I Hybrid Magnet at the University of Nijmegen in Holland.

Peter Wyder, head of the High Field Magnet Laboratory in Grenoble, points out that two recent Nobel prizes were awarded for work in semiconductor physics that depended on the use of high magnetic fields.

Klaus von Klitzing won in 1985 for discovering the Hall effect, the quantum version of the lateral deflection of moving charge carriers in a magnetic field. And Horst Störmer won this year for discovering that this effect can be seen in charge carriers apparently bearing charges a fraction of that of an electron. Both did some of their work at the Grenoble facility.

Wyder is worried that ambivalence towards developing high-field magnets could threaten the future of the Grenoble laboratory, which is financed by France's Centre National de la Recherche Scientifique

(CNRS) and the German Max Planck Society (MPS).

The MPS is uncertain whether to continue its support when the current agreement runs out in 2004, although it might be prepared to do so if the facility increased the number of international partners.

"There is no big demand for these large fields among scientists in our institutes, as far as we can see," says Gerhard Wegner, MPS vice-president for physics and director of the Max Planck Institute for Polymer Research in Mainz. "A new agreement after 2004 would require us to co-finance a major investment of around ECU100 million to modernize the facility."

The MPS is asking its researchers whether such an investment is a priority. It is also seeking opinions on whether the society should support the proposed 100-tesla facility.

The Engineering and Physical Sciences Research Council is consulting UK scientists about future investment in high-field magnets. EPSRC spokesman Neil Williams expects British physicists to ask the research council to participate in either the 100-tesla project or a European high-field continuous magnet facility.

CNRS, which runs a pulsed high-field magnet facility in Toulouse as well as the Grenoble lab, wants to continue development on both fronts. "We will need partners of course," says François Gautier, professor of physics at the University of Strasbourg and deputy director of the CNRS department of physics and mathematics. **Alison Abbott**

German standstill as fewer chairs available

[MUNICH] Contrary to predictions, the number of academic chairs falling vacant in Germany has dropped by nearly a half since 1993, according to a study by the German Association of Universities.

The Wissenschaftsrat, Germany's science council, recently forecast a "high demand" for new professors because the large number hired in the 1970s would reach retirement *en bloc* by the end of this decade.

But the number of advertisements for professorships in biology, medicine, engineering, languages and economics has decreased dramatically. Although the natural sciences and informatics have suffered less than the average, according to the association's survey, advertised positions have still decreased by one third.

The total number of professors in Germany has slightly increased over the past six years, particularly in the eastern *Länder* (states). But the fall in the number of advertised posts could indicate a future decline in the overall number of chairs.

"We are facing the start of a structural change caused by general budget cutbacks," says Hartmut Schiedermair, president of the association. Saving money by not giving opportunities to the next generation is a false economy, he says. "Our young scientists are being left out in the cold."

The number of young scientists gaining the *Habilitation* qualification, which is required by most universities for tenured teaching jobs, has increased by 60 per cent over the past decade. In 1997, a record 1,740 young academics 'habilitated', one third in medicine and a quarter in mathematics and natural sciences. Along with the backlog of academics who habilitated earlier, they had to compete for only 1,112 vacant posts.

The Wissenschaftsrat points out that, although the number of university students is set to fall, the professor-student ratio will fall more sharply if the trend identified by the German Association of Universities continues. Students will end up being short-changed, it warns. **Tilman Kiessling**

Italy backs Third World science body ...

[TRIESTE] The Italian government announced last week that it has agreed to make a long-term financial commitment to the Third World Academy of Sciences (TWAS), the body set up in 1983 at the initiative of the late physicist and Nobel laureate Abdus Salam.

According to academy officials, the move should allow the academy to significantly increase its operating budget within the next few years, boosting its efforts to stimulate scientific excellence in Third World countries.

The academy was set up by Salam alongside the International Centre for Theoretical Physics (ICTP) in Trieste, of which he was director from its foundation in 1964 until 1993. From an initial membership of 44, it

now has more than 400 full members, each recognized for scientific achievement.

Until now, the operating costs of the academy have been largely covered by an annual grant from the Italian government. But the value of this has dropped considerably in recent years, and Italy's financial instability has also created problems.

Under an agreement signed last week, however, the government has promised to present a bill to parliament committing it to increase its payments to the academy from 1.9 billion lire (US\$1.16 million) next year to 3 billion lire in three years' time, and to maintain it at that level thereafter, subject to a biannual review.

This will provide the academy with the

same form of guaranteed income as that already given to two other Trieste institutions: the ICTP and the International Centre for Genetic Engineering and Biotechnology.

"The idea launched by Salam 15 years ago has paid off in many ways in boosting the growth of scientific knowledge in the developing world, and Italy is proud of having been able to contribute to its success," says Facio Bonetti, head of the scientific and cultural division of the Italian foreign ministry.

According to Mohamed Hassan, the executive director of the academy since its creation, Italy's decision to formalize its support was partly prompted by the academy's success in raising money for an endowment fund from developing countries.

The fund, which was launched in 1993, is now more than one-third of the way towards reaching its target of \$10 million. Individual sums of \$500,000 have each been committed by Brazil, China, India, Nigeria and Kuwait, and other countries, including Egypt, Colombia and Senegal, have contributed smaller sums.

"The Italian government was able to see that the developing countries themselves are now committed to our activities," says Hassan. Eventually it is hoped that the academy will be able to use the income from its endowment fund to cover its running costs, allowing all other forms of income to go directly towards projects and other grants.

One of the main contributors towards these at present is the Swedish International Development Agency, which is currently committed to providing \$200,000 a year over three years through its technical assistance division, SAREC, towards a project for training women researchers in sub-Saharan Africa (see box).

Another scheme currently waiting in the wings, says Hassan, is the creation of networks of centres of research excellence in Third World countries. One possible network, involving institutions working in dry-land biodiversity, has recently been submitted for funding to the World Bank's Global Environment Facility.

José Israel Vargas, Brazil's minister of science and the president of TWAS, argues that its expanding activities since its foundation mean that the academy now finds itself "within sight of the vision presented by Salam at the academy's first meeting here in Trieste in 1985".

He adds: "With our operational budget about to be placed on solid ground, our long-term activities gaining in strength through contributions to the endowment fund, and the continued support of outside funding agencies, these are exciting times for TWAS."

David Dickson

... as African women get research career boost

[TRIESTE] The Third World Academy of Sciences (see above) has awarded 25 PhD training fellowships to women in sub-Saharan Africa to study in centres of research excellence in other developing countries.

The novel programme, launched with the support of the Swedish International Development Authority (SIDA), is intended to help boost the participation of women in developing countries in science and technology. All the women, who are under the age of 30, are due to start their fellowships in the next couple of months.

"There are many fewer women than men in the sciences [in the developing world]," says Lydia Makhubu, president of the Third World Organization for Women in Science. Makhubu's organization is an independent, non-profit body, hosted by the Third World Academy of Sciences, which aims to strengthen the participation of women in science-based development and decision-making.

"The aim of the programme is to develop a well trained cadre of women in science and technology – leading to an increase in participation of women in science in these countries."



Women's work: researchers will train in other countries.

Makhubu, a chemist who is vice-chancellor of the University of Swaziland, says one advantage of training women within developing countries – rather than sending them to laboratories in industrialized countries – is that their experience gives them a better chance of finding work afterwards. They also "become more keenly aware of the problems in developing countries that need scientific solutions".

Candidates can choose to do some of their research training in their own countries, something that is crucial for those women who have homes and families they wish to return to. Areas of study will be primarily in the basic sciences.

If the programme is a success, the academy hopes to expand it with additional support from SIDA's Department for Research

Cooperation (SAREC). At present, the aid agency is providing US\$200,000 a year for the three-year programme.

A major factor behind SIDA's sponsorship of the programme was the Swedish government's recent introduction of a new development goal for the agency. According to SAREC official Afzal Sher, one of six development goals approved by the government a year ago "was that SAREC should work towards more gender equality".

"We are aware that the situation with regard to women in science in developing countries is very bad, so it's good to give special support earmarked for women," says Sher. "We have started in a relatively modest way, primarily for sub-Saharan African countries. This may be extended if the pilot works."

It will be several years before SAREC knows the quality of the candidates selected to work on the scheme, and it will take six years to find out whether the programme overall is working. But Makhubu is optimistic. "If you select 25 out of 110 applicants, then you have only the best and most highly motivated," she says.

Natasha Loder

Israeli industry asks for boost to R&D

Haim Watzman

High-technology industry makes up most of Israel's industrial exports. But leading companies, worried by the government's failure to invest in research and development, are threatening to take this work abroad.

[JERUSALEM] The Israeli government is coming under increasing pressure from executives of high-technology companies to do more to promote the growth of their sector of industry, and to provide greater support for research.

Some of these companies, which play a key role in Israel's economy, are warning that they may be forced to transfer their research activities abroad unless they receive more government support. They say it will be cheaper to do the work in the United States.

The Knesset, Israel's parliament, is still considering budget legislation for 1999, but the outlines of the research and development budget are already fairly clear. Despite additional funding, at US\$407 million the budget falls nearly \$200 million short of what the Office of the Chief Scientist says is necessary to fund all worthy projects.

Growth vital to the economy

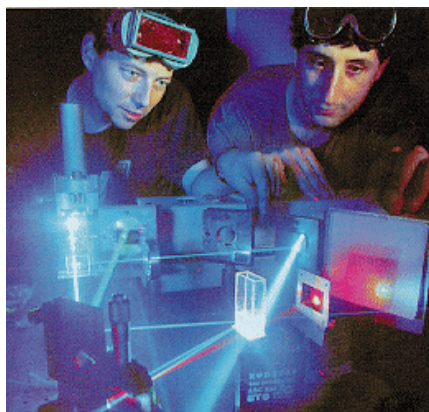
Large companies will be especially hard-hit, says Benny Sofer, deputy to the Chief Scientist. To channel more funds to smaller, newer firms, the finance ministry has set a cap on grants to large companies. Sofer fears that the larger firms will reduce their research and development (R&D) activities as a result.

High-technology products made up 62 per cent of Israel's industrial exports in 1997 — up from 51 per cent less than ten years previously — and were valued at \$9.2 billion. A recent, unpublished survey is said to have found that the software industry alone achieved exports worth \$1 billion in 1997.

Ephie Segal, chief economist in the Office of the Chief Scientist in the Ministry of Commerce and Industry, says that, given Israel's small size, the continued growth of high-technology industries — including electronics, software, pharmaceuticals, biotechnology and aerospace — is vital to the economy.

But he also points out that the future success of such industries depends on current investment in R&D. Israel's investment in R&D of 2.2 per cent of gross domestic product (in 1995) puts it comfortably among the leading industrialized nations. But the Chief Scientist's office and industry minister Natan Scharansky have been pushing for an increase in government R&D funds.

Shuky Abramovich, head of the economics division of the Electronics and Software



In demand: companies face a 'bottleneck' in the supply of trained electrical engineers.

Branch of the Manufacturers' Association, says industry needs another \$210 million to be invested in R&D. "At present, only 55 per cent of requests for grants are approved, and the demand is growing because industry — especially start-up companies — depends on these funds," says Abramovich. "But we don't see any long-term solution to this problem."

According to Abramovich, one temporary solution being considered by the industry ministry is for large companies to be offered the opportunity to pay off in advance — and at a discount — royalties owed to the R&D fund from profits earned from products made with the help of government grants. These funds would be funnelled back into R&D grants during the coming year.

Marketing and staffing problems

The ministry has also asked for government funds to support the industry's marketing activities. Many companies with good ideas have failed to market their products abroad, eventually either failing or being bought by foreign companies which often transfer production and marketing overseas.

Scharansky has been lobbying Prime Minister Binyamin Netanyahu for greater assistance, pointing to the companies' threat to move research activities abroad. Although different high-tech companies have varying needs, most agree that the tax structure creates disincentives for the development and marketing of high-tech products.

There is also a broad consensus that the

country's technical infrastructure — such as telephone lines and charges — is not modernizing quickly enough. And industry says the government is not investing enough in technical and scientific education.

Other concerns are more sector-specific. The largest high-tech sector, electronics and software, suffers from a lack of skilled staff. According to the Manufacturers' Association, the industry is likely to face a shortage of 10,000 engineers over the next five years. Although this reflects a global shortage in these fields, it also threatens to eliminate one of Israel's competitive advantages.

"Until five years ago, manpower was relatively cheap here and Israeli entrepreneurs could take off," says Yoel Raban, an economist at Tel Aviv University who has worked as a consultant for the Industrialists' Association. "But high manpower costs are now a problem for start-up companies."

In response to such concerns, the higher-education system recently announced plans to double the number of Israelis holding degrees in electrical and computer engineering and computer science over the next five years. But industry says this is not enough.

"We're facing a bottleneck and I'm not sure that there is a lot we can do about it in the short run. We seem to have reached the limit," says Raban, adding that it may be necessary to import professionals from countries such as India.

Segal advocates bringing in engineers from the Far East, as is done in the United States. But that would require a major shift in Israeli culture. Until now, universities and industries searching for overseas talent have usually focused on Jewish candidates.

According to Abramovich, the authorities are likely to rectify some of the disincentives for high-tech companies in the tax structure. Changes could include making it easier to give employees stock options, and not requiring the immediate payment of tax when companies merge or split.

In contrast to electronics, biotechnology, the second most important high-technology sector, has a glut of trained professionals. But it lacks investment funds. This problem was highlighted by a survey prepared this year by the Interdisciplinary Center for Technological Analysis and Forecasting at Tel Aviv University (see *Nature* 392, 117; 1998).

Biotechnology had sales of \$336 million in 1997 and employed 1,259 scientists and engineers. But promised government money to encourage investment has not materialized, to the disappointment of industry.

Israel's high-tech leaders are also worried about the effect of the economic crisis in the Far East. According to the daily newspaper *Ha'aretz*, Israeli high-tech companies have raised nearly \$3 billion on Wall Street since the start of 1995, along with another \$1 billion in the past two years by venture-capital funds in New York. □

California's teaching assistants suspend strike and agree to talk

[LONDON] A planned strike by teaching assistants at all nine campuses of the University of California was averted last week when state legislators brokered a 45-day cooling-off period for talks between representatives of the 6,700 assistants and university authorities.

But the assistants say they are still prepared to strike unless talks planned to start this week result in recognition of their membership of a labour union.

The assistants, who are mainly graduate students with some teaching responsibilities, are demanding the right to collective bargaining for pay and conditions. The university has so far refused to accept this, claiming that teaching assistants are students and not employees. Teaching assistants, however, claim that some professors get them to teach core undergraduate courses.

South African minister hits out at former official

[CAPE TOWN] South Africa's Minister of Arts, Culture, Science and Technology, Lionel Mtshali, has accused his departed director-

general, Roger Jardine, of belonging to a "cabal lurking in the woods" that is attempting to impose the African National Congress' agenda on the arts and culture (see *Nature* 396, 298; 1998).

In a lengthy statement, Mtshali accused the cabal of trying to hold the National Heritage Bill hostage, and denied responsibility for the breakdown in relations between Jardine and one of the minister's two deputies, Musa Xulu. Mtshali also vehemently denied accusations that schoolteachers in the former Kwazulu homeland (of which he was education minister) were required to belong to the Inkatha Freedom Party.

Particle detector goes on line soon in Japan

[TOKYO] Belle, a detector intended to explain the mysterious imbalance between matter and antimatter in the Universe, is nearing completion at Japan's High Energy Accelerator Research Organization (KEK) in Tsukuba, Ibaragi Prefecture.

Work began in 1994 on the precision particle detector to study B-mesons produced by KEK's new electron-positron collider, called KEK B-factory. Belle will detect and measure the decay vertices of B-mesons, as well as the energies and momenta

of their daughter particles, with high precision. The detector is scheduled to be moved to the collision point of the collider in February 1999.

Animal rights protester ends hunger strike

[MUNICH] Imprisoned British animal rights activist Barry Horne has ended a 68-day hunger strike intended to force the British government into concessions on the use of animals in research. Horne stopped his strike in a York hospital and accepted treatment last Sunday after receiving a new proposal from the government.

The plan involves the government's Animal Procedures Committee becoming more independent, and collaborating more closely with the Associate Parliamentary Group for Animal Welfare, which includes members of animal rights organizations. Militant activists had threatened to kill ten British scientists involved in animal research if Horne died (see *Nature* 396, 505; 1998).

German X-ray satellite returns its last data

[WASHINGTON] The long-lived German ROSAT X-ray satellite made its final observations last week, after eight years of

studying high-energy emissions in space. ROSAT's High Resolution Imager, one of three main instruments, was irreversibly damaged in September when it scanned too close to the Sun (see *Nature* **395**, 826; 1998).

Mission managers used the imager's remaining supply of xenon gas to reactivate another instrument, the Position Sensitive Proportional Counter (PSPC), which had exhausted its own gas supply in 1994. During a final two days of observations, the PSPC returned data on Supernova 1987a and other selected astrophysical targets. Launched in 1990, ROSAT was designed to last only 20 months.

UK power from biomass helps meet carbon target

[LONDON] The country that gave cricket to the world has scored another first for the use of the fast-growing willow tree, a variety of which is used to make cricket bats — it is using it to produce electricity. Last week John Battle, the energy and industry minister, laid the foundation stone for Britain's first 'significant' power station to be fuelled by biomass.

The power station, to be built in Yorkshire, will turn 42,000 tonnes a year of specially grown willow, as well as agricultural and forestry wastes, into

8 megawatts of electricity, sufficient to heat 8,000 homes. Battle said that, in addition to contributing to sustainable development by using a renewable resource, the station would help to meet national and international carbon reduction targets.

Europe's researchers get high-speed network

[LONDON] Research organizations in countries in northern Europe last week logged into an exclusive high-speed Internet ring that can carry 155 megabits of information per second.

The network, known as TEN-155, will eventually connect research organizations in 16 European countries. Countries in southern and central Europe are expected to join next year. The project has been partly funded by the European Commission as part of the fourth Framework programme of research.

Magnetoresistance attracts German prize

[MUNICH] During a television spectacular last week, Germany's president, Roman Herzog, awarded the new *Deutsche Zukunftspreis* (the 'German Future Prize') to Peter Grünberg, a scientist at the Institute

of Solid State Physics in Jülich, for his discovery of 'giant magnetoresistance'. This property can be used to help develop sensors that can read very compact data on computer hard disks.

In an attempt both to show the 'fun' side of science to the public and to stress 'the importance of innovative research', the prize, worth DM500,000 (US\$300,000), was awarded at the German Theatre in Berlin amid musical and acrobatic displays. Grünberg's selection from a shortlist of four was dramatically revealed when the winner's envelope was opened on stage.

Correction

A recent article on neutrino studies being carried out with the SuperKamiokande detector at the Kamioka Observatory, Institute for Cosmic Ray Research, University of Tokyo, stated that Monbusho has been asked for a budget increase from ¥5 million to ¥10.4 billion for a project to detect neutrinos (see *Nature* **395**, 107; 1998). The second figure should have read ¥1.04 billion.

The article was also incorrect in stating that the Long Baselines Neutrino Oscillation Experiment will send neutrinos through a 250-kilometre underground tunnel; neutrinos can pass through the Earth without the help of a tunnel.

Protest at Nobel omission of Moncada

Sir — The University of El Salvador and its Faculty of Medicine and the National University of Honduras and its Faculty of Medical Sciences wish to express their deep regret and strong protest at the exclusion of Salvador Moncada from the 1998 Nobel Prize in Physiology or Medicine¹.

The first demonstration that nitric oxide is a biological mediator in the cardiovascular system was made by Moncada and his group in *Nature*² in 1987. This paper provided the first direct and unequivocal evidence for the hypothesis that endothelium-derived relaxing factor (EDRF) might be nitric oxide or a related substance. Moncada also showed that the biosynthetic pathway for the generation of nitric oxide in the cardiovascular system was from the amino acid L-arginine³. Without these two seminal papers the field of nitric-oxide research would not exist. Moncada went on to make some of the most significant contributions to the field.

To distinguish earlier research, which can only be deemed part of this field with hindsight and in the light of Moncada's work, without recognizing his work, is tantamount to an attempt to rewrite the history of this discovery.

We would like to believe that this is only the result of a regrettable mistake. We expect that the Nobel committee will repair the damage to its credibility.

**José Benjamín López Guillén,
Eduardo A. Espinoza Fiallos**

*Universidad de El Salvador, Ciudad Universitaria,
Final 25 Avenida Norte, San Salvador, El Salvador*
Ana Belén Castillo, Gustavo A. Vallejo
*Universidad Nacional de Honduras,
Ciudad Universitaria, Boulevard Suyapa,
Tegucigalpa, Honduras*

Sir — The Nobel committee's decision to award the prize in physiology or medicine to Robert Furchgott, Louis Ignarro and Ferid Murad has aroused controversy. I think it important to point out that, although Ignarro's 1987 paper in the *Proceedings of the National Academy of Sciences*⁴ is more widely recognized, in fact he submitted in 1986 the first publication to conclude that Furchgott's endothelium-derived relaxing factor is nitric oxide (*Circulation Research*⁵).

Although I do not want to quibble about the one month's difference in submission date between this and Moncada's *Nature* paper², which appeared in print before Ignarro's paper, it is inaccurate to conclude that Ignarro's findings came six months after Moncada had already made this conclusion, as stated in your report¹.

Jack R. Lancaster Jr

*Louisiana State University Medical Center,
1901 Perdido St, New Orleans, Louisiana 70112, USA*

Sir — Your article describing this year's prize in physiology or medicine starts "The Nobel committee has once again sparked controversy..." as if there were one Nobel committee¹. There are, in fact, five committees, one for each prize — physics, chemistry, physiology or medicine, literature, and peace. Another common misunderstanding is that the prize for economics is not a Nobel prize, but is donated by the Bank of Sweden in memory of Alfred Nobel.

The Nobel Foundation has nothing to do with the selection process but just administers the funds. The prize-awarding institutions and their committees are autonomous. They are the Swedish Academy

of Sciences for chemistry and physics; the Nobel Assembly of the Karolinska Institute for physiology or medicine; the Swedish Academy for literature; and the Norwegian Parliament for peace.

Bo G. Malmström

*Department of Chemistry, Göteborg University,
SE-405 30 Göteborg, Sweden*

Our article¹ was written in the belief that *Nature* readers are aware that there are several committees, but we are glad to clarify matters. The most troublesome aspect of the Nobel process is the apparently unchangeable fact that the prizes are distributed according to the terms of Nobel's will, which states that the number of recipients in each category shall be limited to three. We have received several letters protesting at the fact that Moncada was not awarded a prize. In relation to Lancaster's point, the *Circulation Research* paper⁵ appeared in December 1987, six months after Moncada's paper in *Nature*². In that paper Ignarro and colleagues tentatively concluded that "EDRF from artery and vein is either nitric oxide (NO) or a chemically related radical species". The earlier paper from Moncada's group was also cautious in suggesting the identity of EDRF and NO on the basis that the release of NO was sufficient to account for the former substance's biological activity. —
Editor, *Nature*

1. Howlett, R. *Nature* **395**, 625–626 (1998).

2. Palmer, R. M. J., Ferrige, A. G. & Moncada, S. *Nature* **327**, 524–526 (1987).

3. Palmer, R. M. J., Ashton, D. S. & Moncada, S. *Nature* **333**, 664–666 (1988).

4. Ignarro, L. J. et al. *Proc. Natl Acad. Sci. USA* **84**, 9265–9269 (1987).

5. Ignarro, L. J. et al. *Circ. Res.* **61**, 866–879 (1987).

Bitter pill to swallow over medical education

Sir — Three years after a review committee evaluated biomedical research in Austria, another external expert committee has presented equally disastrous findings on the quality of Austrian medical training (*Nature* **377**, 468; 1995 & **395**, 832; 1998).

In 1995 a research evaluation committee organized by the European Molecular Biology Organization strongly urged certain improvements. No action was taken. Indeed, the opposite of the recommended changes has been achieved. Lifetime tenure is now being granted automatically to all assistant professors, who are required to teach; institutes are merged without

evaluation; and none of the assessments of groups or institutes has led to any change in their levels of funding.

Perhaps our authorities consider that external advice (after objective evaluation) is unnecessary. The failure to act upon the previous committee's recommendations suggests that the latest report, a Dutch committee's assessment of university medical education in Austria, will be met with the same attitude of 'intelligent neglect'.

As professors of pharmacology in Austria, with considerable experience of teaching in other countries, we feel that the problems facing medical training in Austria and the consequences for other European countries demand greater coverage.

Access to medical study is unlimited, examiners can be chosen freely by students

from any of the three medical faculties, examinations can be repeated up to four times, and there are few compulsory lectures. Many teaching staff lack an MD qualification and do not know the basic requirements of clinical work. Many students qualify — on average after nine years — without having been trained to solve even the simplest clinical problem.

Responsible European Union committees must analyse this unsatisfactory situation because Austrian doctors will be able to work freely in any other country of the union.

Hartmut H. Glossmann

*Institute of Biochemical Pharmacology,
University of Innsbruck, 6020 Innsbruck, Austria*

Bernhard A. Peskar

*Institute of Experimental and Clinical Pharmacology,
University of Graz, 8010 Graz, Austria*

Benchmarking international research

The United States leads the world in the impact its research makes in almost every scientific domain. But European countries could challenge that if they were to divide their labours according to their strengths.

Jonathan Adams

In 1997 the Higher Education Funding Council for England, which funds research and teaching in more than 100 universities and other institutions, decided to develop measures for the international standing of, and national needs for, different subject areas. One of the quantitative analyses it used was our study on bibliometric performance indicators¹. This article uses that study (which also contains references to relevant previous studies and a detailed discussion of assumptions and methods) to consider more broadly the international competitiveness and management of research.

May² looked at various countries' shares of world outputs (papers published) and citations. He noted that research performance can be measured in different ways, and should take into account such factors as population size and national investment in research. May subsequently extended his analysis to consider the funding patterns that underpinned these differences in performance³. There are very marked, sometimes more than two-fold, differences in the productivity (outputs per unit funding) of some countries, such as the United States and United Kingdom, compared to others, such as France, Germany and Japan.

Our study categorizes research in a different way to May, and thereby provides a cross-check on his results. Our sponsor was interested in identifying relative strengths and weaknesses in national performance between fields. We were also interested in a key point raised in May's study: that a large part of the difference in national performance, absolutely and relative to resources, arises from differences in institutional settings and hence management cultures. For this report, we have looked at total country performance. The focus is an international comparison for several countries (England — because the study was sponsored by the funding council for England — with Australia, Canada, France, Germany, Japan and the United States) against a world baseline.

The United States has the greatest output and citation share and has a leading international impact. England and Canada also perform well across most fields. The output of France, Germany and Japan is significant, but their research impact is less strong and is also more evidently focused in particular fields. The results reflect a pattern of association between research competitiveness and the structure of research management. Within Europe, there is complementarity in research strengths with a bioscience lead in

Table 1 Impact measures rebased against world baseline

UOA code	Unit of assessment	US	England	Canada	France	Germany	Australia	Japan
1	Clinical Laboratory Sciences	1.34	1.06	1.02	0.91	0.94	0.92	0.74
2	Community-based Clinical Subjects	1.35	1.11	0.99	0.72	0.71	0.79	0.75
3	Hospital-Based Clinical Subjects	1.34	1.10	1.12	0.81	0.77	0.95	0.75
4	Clinical Dentistry	1.12	0.90	1.24	0.53	0.76	0.55	0.86
5	Preclinical Studies	1.41	1.21	0.63	0.69	0.94	0.63	0.53
6	Anatomy	1.41	1.21	0.63	0.69	0.94	0.63	0.53
7	Physiology	1.29	1.13	0.94	0.72	1.06	0.78	0.80
8	Pharmacology & 9 Pharmacy	1.29	1.67	1.04	0.99	0.87	0.97	0.72
10	Nursing	1.20	0.63	0.77	0.67	0.41	0.60	0.99
11	Other Studies and Professions Allied to Medicine	1.44	1.05	0.89	0.47	0.73	0.63	0.71
12	Biochemistry	1.35	1.12	0.94	0.88	0.98	0.82	0.76
13	Psychology	1.14	1.13	1.02	1.05	0.86	0.78	0.89
14	Biological Sciences	1.43	1.19	0.89	0.95	1.08	0.82	0.82
15	Agriculture	1.35	1.50	1.18	1.16	1.21	1.25	0.88
16	Food Science and Technology	1.39	1.25	1.28	1.05	0.62	0.93	0.85
17	Veterinary Science	1.39	1.48	1.07	1.07	0.66	0.94	0.99
18	Chemistry	1.63	1.09	1.23	0.99	1.05	0.99	0.94
19	Physics	1.54	1.21	1.23	1.09	1.16	1.03	0.88
20	Earth Sciences	1.45	1.14	0.99	1.05	1.04	1.11	0.78
21	Environmental Sciences	1.32	1.11	1.20	1.00	1.00	1.11	0.61
22	Pure Mathematics	1.28	1.41	1.02	0.97	0.91	1.12	0.81
23	Applied Mathematics	1.57	1.14	1.08	1.32	1.33	0.83	0.82
24	Statistics and Operational Research	1.35	1.29	1.09	0.82	0.83	1.24	0.77
25	Computer Science	1.34	0.86	1.12	0.93	0.79	0.93	0.57
26	General Engineering	1.39	0.96	0.94	1.07	1.00	0.85	0.95
27	Chemical Engineering	1.46	0.97	1.30	1.34	0.75	1.20	0.91
28	Civil Engineering	1.25	0.94	0.96	1.16	0.95	1.31	0.66
29	Electrical and Electronic Eng.	1.42	0.99	0.99	1.08	1.17	0.87	0.99
30	Mechanical, Aeronautical and Manufacturing Engineering	1.35	1.03	0.96	1.08	0.93	1.02	0.87
31	Mineral and Mining Eng.	1.13	1.31	1.16	1.47	1.19	1.49	0.91
32	Metallurgy and Materials	1.60	1.07	1.02	1.05	1.04	0.91	0.94
33	Built Environment	1.41	0.71	1.13	0.95	0.65	0.93	0.76
34	Town and Country Planning	1.16	1.16	1.05	0.81	0.56	1.02	0.47
35	Geography	1.32	0.99	1.07	1.17	1.07	1.11	0.75
36	Law	1.08	0.35	0.73	0.27	0.21	0.52	0.47
37	Anthropology	1.39	0.96	0.83	0.37	0.49	0.75	0.50
38	Economics and Econometrics	1.24	0.86	0.88	0.70	0.62	0.56	0.55
39	Politics and International Studies	1.35	0.75	0.84	0.31	0.38	0.52	0.45
40	Social Policy and Admin.	1.08	1.18	1.16	0.41	0.54	0.90	0.37
41	Social Work							
42	Sociology	1.39	0.96	0.83	0.37	0.49	0.75	0.50
43	Business and Mgmt Studies	1.21	0.77	1.01	0.74	0.64	0.72	0.66
44	Accountancy	1.24	0.85	0.89	0.72	0.63	0.57	0.55
61	Library and Info. Management	1.11	1.05	1.23	0.91	0.53	0.85	0.22
65	Communication, Cultural and Media Studies	1.07	0.76	1.18	0.77	1.01	0.87	0.25
68	Education	1.17	0.65	0.97	0.56	0.36	0.91	0.52
	Mean	1.32	1.07	1.02	0.86	0.82	0.89	0.70
	Variance	0.02	0.07	0.03	0.08	0.07	0.05	0.04

England and a physical science/engineering lead in Germany and France. The close link between basic biomedical research and its development may play to the strengths of the United Kingdom's university-based research

system, while success in technology-linked research requires well managed industrial development that builds on an institute-based system. Research impact declines in fields with high output in Australia, Canada

and England: this may reflect the limits to worthwhile investment.

Data sources and mapping

The main source of information for our research was the database of journals of the international Institute for Scientific Information (ISI). These journals are divided into 94 more or less discrete categories. Within each category we can add up the number of papers that originated from a given country over a period, and compare this with the total for that journal set. This gives us a measure of the relative output of each nation for that category. We can then take these papers and compute the number of citations they received. The average number of citations per paper, within this ISI subject area, can then also be compared internationally.

However, the ISI categorization did not meet our needs as it stood. Our sponsor wanted comparisons based on its own 69 disciplinary Units of Assessment (UOAs). We therefore needed to link ISI categories to each UOA in the most informative way possible. The initial seed for our research was the database of publications submitted by researchers in English universities to the 1996 Research Assessment Exercise (RAE). Every active member of academic research staff, associated with a given UOA, is required to submit up to four publications of international quality for assessment. The journals in which these publications appear are thereby associated with a given UOA, creating a categorization of journals determined by researchers themselves.

We took 47 of the 69 disciplinary UOAs (the rest were mainly humanities subjects) and compared the journals in each UOA with the lists created by the ISI. Unlike the ISI's discrete categories, a journal might appear several times under different UOA heads.

About 80% of the RAE submissions appeared in journals that were on the ISI lists. So, for a given UOA, we could establish which ISI categories matched it most closely. For each of the 47 UOAs, we drew down a set of publications and citations from the corresponding ISI categories for 1988–96. This period gave a reasonably up-to-date picture of research activity that could be benchmarked against other indicators¹. A Table in our report¹ gives a complete list of the ISI categories we assigned to each UOA.

We grouped data at two levels, corresponding to the 47 UOA disciplinary units, such as physics, and to amalgamations of these units ('Super-UOAs') as 13 major subject groups, such as physical sciences (Table 2). As data are aggregated some detail is obscured, but general patterns also emerge⁴.

The overall picture emerging from our study is consistent across the analyses based on the RAE-derived journal set, the journal sets for UOAs created from ISI categories, and the supra-disciplinary groupings of SUOAs. We therefore have some confidence

that this approach has been meaningful both in terms of establishing a benchmark and in extending this to international comparison.

International comparisons

Our 1988–96 dataset, based on the ISI categories grouped to the 47 UOAs, produced impact measures rebased to world average (Relative Citation Index, RCI)¹. The RCI data are shown by country (Table 1). Super-UOA performance is shown by country (Table 2). We ranked UOAs separately for each country in order of performance relative to world average (Table 1). The upper quartile of each country's 'league table', identifying its areas of core strength internationally, is collated (Table 3).

United States

The United States has a high and consistent (low variance) impact across UOAs. Its average impact (1.32) and international ranking (1.38) put it clearly ahead of other countries and it ranked first in 35 of the 47 UOAs and second in nine others. US performance is partly a consequence of a high volume, which has a significant effect on world averages.

- US output share is greater than 40% and citation share is greater than 50% in all but two UOAs.
- In the UOA league (Table 3) US performance in physical sciences and applied maths is high, creating an outstanding world lead.
- US engineering is bimodal with electrical, chemical and mechanical engineering at the top and civil (and related) and mining engineering much lower ranked.
- There is higher variance in performance among the biological and clinical sciences.

Japan

Japan's relative volume (more output than any other country except the United States)

reflects a significant contribution to widely used, international journals. Overall, however, its average rank of 6.17 and average impact of 0.7 compared to the world average suggest that its contributions to basic research have yet to be fully realized.

- Strengths lie in engineering (up to world average and ahead of England and Canada) and in the physical sciences, particularly related to electrical engineering and to materials.
- Japan is weaker in biosciences.
- The contribution made by particular sectors is difficult to pick out from whole-country analyses. The performance of specific institutes may be highly competitive.

England

The average England UOA performance puts it second to the United States. England is ranked second or third in 10 of the 13 SUOAs and performs evenly across subjects (Table 2).

- Strengths are in the core biological and physical sciences UOAs, which dominate the top third of England's league table (Table 3).
- England is strong in biomedical sciences, as noted elsewhere⁵. In the preclinical and biological sciences it is fractionally second behind the United States and far ahead of Germany. This strength is sustained on a relatively small publication share.
- Despite a low output share in the physical sciences, England is essentially part of a second rank with Germany and Canada.
- In engineering there is a consistent dip in citation share¹. England is in a trailing group with Canada and Australia and well behind its European competitors.

Germany

German performance is worse than that of the United States or England. There is concentrated output and impact in core fields.

Table 3 Upper quartile (most highly rated) of units of assessment for each country ranked by citation impact (normalized to world baseline)

Rank	US		England		Canada	
	Unit of assessment	RCI	Unit of assessment	RCI	Unit of assessment	RCI
1	Applied Mathematics	1.62	Pharmacology/Pharmacy	1.67	Chemical Engineering	1.29
2	Physics	1.59	Agriculture	1.44	Food Science and Tech.	1.28
3	Chemistry	1.57	Pure Mathematics	1.41	Clinical Dentistry	1.24
4	Metallurgy and Materials	1.52	Statistics and Oper. Res.	1.36	Library/Information Management	1.23
5	Electrical and Electronic Eng.	1.47	Veterinary Science	1.33	Chemistry	1.19
6	Chemical Engineering	1.46	Mineral and Mining Eng.	1.31	Comm. & Media Studies	1.18
7	Mechanical & Manufact. Eng.	1.45	Food Science and Tech.	1.25	Physics	1.18
8	Earth Sciences	1.44	Anatomy	1.21	Social Policy & Social Work	1.16
9	Professions allied to Medicine	1.44	Pre-Clinical Studies	1.21	Mineral and Mining Eng.	1.16
10	Veterinary Science	1.44	Social Policy & Social Work	1.18	Built Environment	1.13
11	Community Based Clinical	1.43	Applied Mathematics	1.17	Applied Mathematics	1.13
12	Biological Sciences	1.42	Physics	1.17	Environmental Sciences	1.12
13	Anatomy	1.41	Biological Sciences	1.16	Computer Science	1.11

Table 2 Super-UOA performance for comparator countries

Super unit of assessment		US	England	Canada	France	Germany	Australia	Japan
<i>Rebased to world average</i>								
SU01	Clinical Medicine	1.35	1.09	1.10	0.83	0.81	0.94	0.75
SU02	Clinical Dentistry	1.12	0.90	1.24	0.53	0.76	0.55	0.86
SU03	Preclinical Sciences	1.45	1.41	0.93	0.87	0.95	0.83	0.63
SU04	Biological Sciences	1.46	1.24	0.91	0.99	1.02	0.78	0.86
SU05	Mathematics	1.41	1.22	0.99	1.11	1.23	0.91	0.88
SU06	Physical Sciences	1.58	1.14	1.21	1.04	1.10	1.00	0.91
SU07	Engineering	1.34	0.97	0.97	1.15	1.06	0.95	1.00
SU08	Earth & Environment	1.37	1.08	1.08	1.09	1.05	1.12	0.71
SU09	Social, Health & Community	1.32	1.10	0.98	0.74	0.73	0.78	0.76
SU10	Social Sciences, incl. Law & Politics	1.25	0.81	0.85	0.36	0.37	0.62	0.50
SU11	Business & Management	1.23	0.83	0.92	0.72	0.63	0.60	0.58
SU12	Psychology	1.14	1.13	1.03	1.03	0.87	0.78	0.89
SU13	Other	1.15	0.74	1.05	0.71	0.45	0.94	0.32
Average impact		1.32	1.05	1.02	0.86	0.85	0.83	0.74
<i>Rank order</i>								
SU01	Clinical Medicine	1	3	2	5	6	4	7
SU02	Clinical Dentistry	2	3	1	7	5	6	4
SU03	Preclinical Sciences	1	2	4	5	3	6	7
SU04	Biological Sciences	1	2	5	4	3	7	6
SU05	Mathematics	1	3	5	4	2	6	7
SU06	Physical Sciences	1	3	2	5	4	6	7
SU07	Engineering	1	6	5	2	3	7	4
SU08	Earth & Environment	1	5	4	3	6	2	7
SU09	Social, Health & Community	1	2	3	6	7	4	5
SU10	Social Sciences, incl. Law & Politics	1	3	2	7	6	4	5
SU11	Business & Management	1	3	2	4	5	6	7
SU12	Psychology	1	2	4	3	6	7	5
SU13	Other	1	4	2	5	6	3	7
Average rank score		1.08	3.15	3.15	4.62	4.77	5.23	6.00

German scientific research may be more concentrated in managed programmes than in England. For example, the mission-orientated Max-Planck Institutes would have a significant effect on performance in specific areas. We return to this point in discussion.

- There is particular strength, with high output and impact, across applied mathematics,

physics and electronic engineering (Table 3).

- There is exceptional output and sound impact in metallurgy and materials and in chemistry, but overall research quality in physical sciences is diluted by volume¹.

- Strengths are not always allied to consistency across broader fields. Mechanical and chemical engineering rank sixth and seventh

internationally, so the engineering SUOA slips to third overall (Table 2).

- Biochemistry output is greater than England's but impact is below the world average (Table 1).

France

Despite a relatively modest output volume, France's performance in engineering is notable. If the research agency CNRS directs a more targeted research programme than is typical in universities, there may be a 'management' effect as for Germany.

- Most engineering UOAs are in the upper quartile (Table 3) and all are in the top half.

- Mining and chemical engineering have a high impact on a relatively small output. Civil and mechanical engineering are also strong.

- Engineering strength spreads across related areas, including applied maths and physics.

- Biological subjects — well represented within the CNRS — are weaker. Biochemistry and the clinical sciences also perform poorly, among French UOAs and internationally.

- France is also weak in the health and community area, suggesting that health services may be an area of general under-achievement, although psychology is stronger.

Canada

Canada broadly ranks third behind the United States and England, with an average rank across UOAs of 3.19 and an overall relative impact of 1.02 compared to the world average. There are no peaks or concentrations of excellence, variance in impact is less than for other countries and impact is maintained across a broad front on a low relative volume.

- Canada's average SUOA performance is on a par with that of England (Table 2). It is, however, difficult to pick out areas of consistent strength in depth.

There were 47 UOAs in this analysis; the upper quartile thus includes the top ranked 12 but the marginal 13th has been included for interest.

France		Germany		Australia		Japan		Rank
Unit of assessment	RCI	Unit of assessment	RCI	Unit of assessment	RCI	Unit of assessment	RCI	
Mineral and Mining Eng.	1.47	Applied Mathematics	1.28	Mineral and Mining Eng.	1.49	Electrical and Electronic Eng.	1.00	1
Chemical Engineering	1.42	Agriculture	1.25	Civil Engineering	1.30	Nursing	0.99	2
Applied Mathematics	1.25	Physics	1.20	Chemical Engineering	1.24	Metallurgy and Materials	0.99	3
Civil Engineering	1.16	Electrical and Electronic Eng.	1.19	Statistics and Oper. Res.	1.20	General Engineering	0.95	4
Geography	1.14	Mineral and Mining Eng.	1.19	Agriculture	1.19	Chemical Engineering	0.92	5
Mechanical & Manufact. Eng.	1.12	Metallurgy and Materials	1.11	Pure Mathematics	1.12	Agriculture	0.92	6
Veterinary Science	1.11	Geography	1.06	Geography	1.12	Chemistry	0.91	7
Physics	1.08	Physiology	1.06	Environmental Sciences	1.11	Mineral and Mining Eng.	0.91	8
General Engineering	1.07	Biological Sciences	1.05	Earth Sciences	1.10	Physics	0.91	9
Agriculture	1.06	Chemistry	1.05	Mechanical & Manufact. Eng.	1.08	Veterinary Science	0.91	10
Electrical and Electronic Eng.	1.05	Earth Sciences	1.03	Town and Country Planning	1.02	Psychology	0.88	11
Food Science and Tech.	1.05	Environmental Sciences	1.02	Chemistry	0.97	Clinical Dentistry	0.86	12
Earth Sciences	1.03	Comm. & Media Studies	1.01	Pharmacology & Pharmacy	0.97	Applied Mathematics	0.85	13

● Among UOAs, Canada is generally stronger in physical sciences than biological sciences while engineering tends to be rather mixed in quality (Table 3).

Australia

Australia has a comparable average impact to France and Germany despite an output less than half either of theirs. At an aggregated level, it also has a more consistent performance, like that of Canada.

- Strength in mining and civil engineering is offset by weaker performances in electrical engineering and computer science.
- Performance is good in the Earth and environmental sciences, according with strengths in mining and agriculture. This is perhaps culturally related to a successful history of natural resource exploitation.
- Australia's relatively low volume of research makes further analysis less appropriate.

Structural comparisons

Two key impressions emerge from the international overview. First, competitiveness in research tends to be a package: some nations perform well (United States, England) at both a specific (UOA) and more general (SUOA) level while others perform less well. This impression is supported by these different levels of data grouping and by different indices¹. The necessary — and frequently noted — caveat is that there may be an anglophone bias that tends to disadvantage some countries. Against this, the volumes for France, Germany and Japan relative to England and Canada suggest that this should no longer be a major issue.

The second impression is that some countries perform relatively evenly across areas, while others have more evident and characteristic core strengths. This can be quantified in various ways. For example, there is a widely recognized correlation between publications and citations: the more you publish, the more you get cited (often by yourself). This holds true at different levels of aggregation for disciplines and institutions. Somewhat surprisingly, however, this correlation is much weaker for England than it is for other countries (correlation coefficient across 47 UOAs: England = 0.78, Germany = 0.93, United States = 0.94). In other words, England's citation rates are higher than expected in areas of low relative output. Such differences may suggest that some countries perform better than others, even in areas where their relative investment is less³. Alternatively, other nations may develop a consensus about investment and effort and concentrate talent behind core programmes⁶.

Discussion

Differences, even complementarity, in relative international performance may argue both for increased national selectivity and greater governmental cooperation in the support of science. If so, then this would provide strong support for recent policy initiatives in

this direction in France and Germany and selective funding initiatives in the United Kingdom. Similar initiatives may prove rewarding elsewhere.

Performance in European engineering and biosciences, for example, contrasts significantly. The performance of France, notably in engineering, and Germany, where there is concentrated activity and impact in physical and applied sciences, reflects traditional and core strengths. In England, however, a weak research and development base outside universities and a declining industrial sector may have contributed to its engineering being out-ranked by both these competitors. On the other hand, a high English citation share in biosciences means that a moderate share of world publications has a much higher impact than for Germany or France. This reflects both breadth and diversity in the English research base in this sector, possibly due to the pharmaceutical and biotechnology industries as well as research council and charity-funded units. Hicks and Katz⁷ also noted that UK industrial research in the life sciences is more highly cited than UK research as a whole.

The relative performance of UK engineering and biological sciences could therefore reflect an entirely appropriate shift in the UK industrial base — and national research focus — over the past few decades, rather than representing any malaise or a lack of esteem for applied research. This would suggest that the recent announcement by the UK government of increased university research funding focused on priority areas in biological sciences is astute selectivity. More managed structures, by contrast, have given the French and German systems a strong and consistent lead in physical sciences and in engineering.

It is tempting to argue that the overall research performance (and competitiveness) of the European Union could be enhanced by the United Kingdom 'buying' its engineering from France and Germany in exchange for biotechnology know-how. There is a close link between basic biomedical research and its development, whereas it is argued that success in technology-linked research requires well managed industrial development and innovation quite independent of research creativity. The differences between the UK, French and German research cultures may play to different strengths in these sectors.

Our results show a pattern of association between research competitiveness and the structure of research management. The significance of national patterns of coincidence — or vice versa a negative relationship — between effort (relative output) and quality (relative impact) has been discussed at some length by May^{2,3}. Institutional settings can have a significant impact on differences in performance. May saw the more volatile environment of a university, the normal host for basic research in the United States, England, Canada and Australia, as a better environ-

ment than a dedicated research institute.

We have shown that those research economies that are typically more directed through, for example, well established and mission-led institutes such as the Max-Planck and the CNRS, tend to have positive associations between research output (as a proxy for activity) and the impact of that research. There is a link with applied research areas and some general relationship might also be inferred between historical traditions and well founded industrial sectors. In those economies in which the management of research activity is led by higher education institutions, on the other hand, research performance is good across the broad sweep of UOAs (that is, variance in impact is lower in Australia and Canada than in France and Germany). Performance in core research may be high despite relatively low world share. Canada is third behind England and the United States, and Australia competes with France and Germany. There is a similar pattern of relative research effectiveness between these countries in relation to population and to total and civil research expenditure⁵.

A similar model for international research complementarity and cooperation might be suggested for the Pacific Rim countries. The strengths of the more competitive parts of the research base and the best features of the research management systems of Australia and Japan could fruitfully be coordinated. Australia is extraordinarily competitive on a relatively small activity base, while Japan provides both the resources and infrastructure but has yet to secure the direct benefits in terms of competitive research quality. Working together, these national systems could provide a strong challenge to Europe and to North America.

A final question is why research impact actually declines in areas of high output in Australia, Canada and England. A reasonable presumption might have been that it would at least maintain par. One answer might be that, even if the frontiers of research are endless, each country has only a limited quantum of good research to offer. Investment beyond that point is nugatory; greater quantity is inevitably poorer quality. Perhaps the UK government's selective allocation to priority biosciences and research minister Claude Allègre's aim for higher French impact will provide tests of this idea. □

Jonathan Adams is at the Centre for Policy Studies in Education, University of Leeds, Leeds LS2 9JT, UK.

1. Adams, J. et al. *Benchmarking of the international standing of research in England: report of a consultancy study on bibliometric analysis* (CPSE, Univ. Leeds, 1998) (<http://www.leeds.ac.uk/>).
2. May, R. M. *Science* **275**, 793–796 (1997).
3. May, R. M. *Science* **281**, 49–51 (1998).
4. Clarke, R. & Grahame, J. W. "Analysis of structure in HEFCE RAE bibliometric data". Annex to ref. 1.
5. Office of Science and Technology. *The quality of the UK science base* (Department of Trade and Industry, London, 1997).
6. Hills, G. *Chemistry and Industry* **7**, 284 (1998).
7. Hicks, D. & Katz, S. *The Changing Shape of British Industrial Research*. STEEP Special Report No. 6 (Univ. Sussex, 1997).

C-quest in the Amazon Basin

I. Colin Prentice and Jon Lloyd

Whether Amazonia is a carbon-dioxide source or sink has a significant effect on estimates of the global carbon budget. On an annual basis, El Niño, and its influence on precipitation in the region, seems to be a controlling factor.

What is happening to all of the extra carbon dioxide being pumped into Earth's atmosphere as a result of human activities? Seven peta (10^{15}) grams of carbon are now being emitted annually by fossil-fuel burning, and another 1–2 Pg by forest clearance in the tropics (for crops and pastures, which contain far less carbon). Only about half of this carbon stays in the atmosphere. The oceans absorb a further quarter or so, leaving at least 2 Pg C yr⁻¹ to be accounted for by uptake into terrestrial ecosystems¹. The 'C-quest' for the carbon² presumed to be sequestered on land has been on for two decades — during which time, of course, people have continued to burn fossil fuels and clear forests.

Two petagrams might sound a lot, especially when expressed as 2 billion tonnes. But it is only about 0.1% of the carbon stored in land ecosystems, and so is not easy to find. Interest has centred in particular on the Amazon Basin, which is still the largest area of wet tropical forest on Earth and, on theoretical grounds, a prime candidate as a carbon sink³. Flux measurements have confirmed that there is substantial carbon uptake there⁴ — so much, indeed, that the remaining forests of Amazonia might be gaining as much carbon as is being lost through forest clearance in the same region. This inference finds support in the biogeochemical model results presented by Tian *et al.* on page 664 of this issue⁵, but with an important rider.

Climate variability causes large swings, so that in El Niño years these ecosystems can become a source of carbon release into the atmosphere. We have, then, new and unresolved questions about the future of this 'service' under a changing climate.

The problem with measurements of carbon flux is that they are difficult to make, and we have too few of them. The area represented is about 1 km², and no Amazonian sampling campaign so far has lasted more than one year. Yet the climate varies enormously, both from year to year and spatially, as seen in the variety of vegetation — the basin includes dry savanna (*cerrado*; Fig. 1), permanently wet forest and everything in between.

Tian *et al.*⁵ circumvent this problem by driving their model with month-by-month climate observations, interpolated to a basin-wide grid. After showing that the model can approximate the published (albeit sparse) flux measurements for specific places and times, they use the full model set-up to reach two conclusions. First, the basin was a carbon sink (on average) during the period 1980–94, because of increased photosynthesis resulting from increasing CO₂ concentration^{2,3}. Second, year-to-year variation (due to inhibition of photosynthesis by lack of soil moisture) exceeds the average sink effect, so that the influence of drought outweighs the sink effect in 'strong' El Niño years when precipitation in much of the Amazon Basin is

severely reduced. Such results could be model-dependent, but a run we have performed with a structurally quite different model gives an encouragingly similar conclusion (Fig. 2, overleaf). In Fig. 2, values below the zero line indicate downward fluxes from the atmosphere into the biosphere — that is, when the region was acting as a carbon sink. The peaks, when the region acted as a weak sink or as a source, correspond to the three strongest El Niño events during this period (1982–83, 1986–87, 1991–92).

There is independent, geophysical evidence¹ for carbon sinks in tropical lands. The most comprehensive global calculation to date⁶, seeking the configuration of land and ocean CO₂ fluxes that best accounts for 1980–95 global observations of atmospheric CO₂, the O₂:N₂ ratio and stable carbon isotope signatures, assigned average net fluxes close to zero to tropical land areas. Amazonia was estimated as having a net balance of around 0.1 Pg C yr⁻¹ (that is, natural sinks exceeded deforestation sources by this amount). Tree biomass measurements⁷ on almost 100 sample plots also imply that lowland tropical forests of Amazonia may have taken up as much as 0.4 Pg C yr⁻¹ during 1975–96. Both of these approaches involve big approximations. But they point in the same direction as the models (Fig. 2), both of which estimate an average uptake of 0.2 Pg C yr⁻¹ during 1980–94.

The growth rate of atmospheric CO₂ concentration varies by ± 2 Pg C yr⁻¹, with maximum rates in strong El Niño years². Terrestrial biogeochemical models can explain much of this variation through the effect of El Niño on tropical forests⁸. And few would dispute Tian and colleagues' identification of precipitation (acting through soil moisture) as a key underlying control. The vegetation gradients across the Amazon Basin reflect variation in the length and severity of the dry season, and flux measurements even from the wet central region show reduced photosynthesis during drier months⁹.

The qualitative conclusions of Tian *et al.* should therefore not be controversial. But what about the numbers? As Fig. 2 shows, biogeochemical models differ quantitatively in their simulation of CO₂ and climate effects. They must be better constrained in future by flux measurements made over longer periods and in a wider variety of environments. Models should also be able to reproduce the empirical features of the global atmospheric CO₂ record. For example, we do not yet know whether biogeochemical models can simulate the opposing, time-lagged response of CO₂ growth rate to El Niño¹⁰ — or even what processes lie behind it. Such understanding is crucial for long-term predictions.

Nowadays a lot depends on assessments of terrestrial carbon budgets. For years, this line of research was a source of fascinating geophysical and ecological problems. Now,

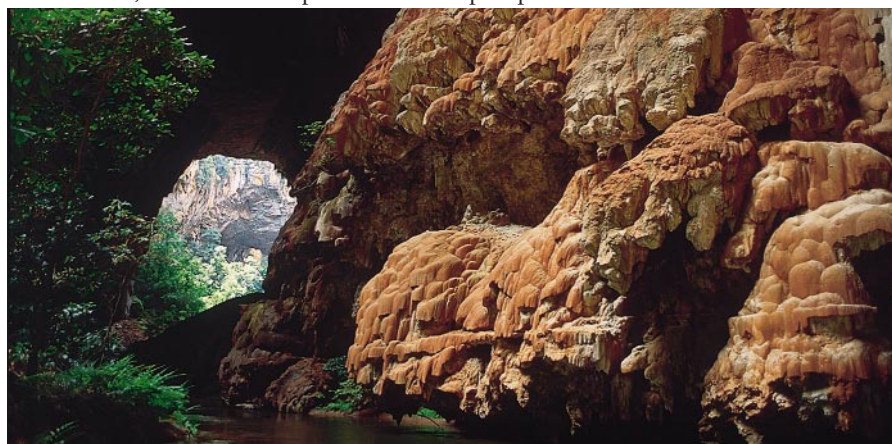


Figure 1 *Cerrado* in Amazonia, where rainfall strongly limits plant growth. Even in the more familiar rainforests of the region, soils become dry enough during part of the year to put a significant constraint on tree productivity.

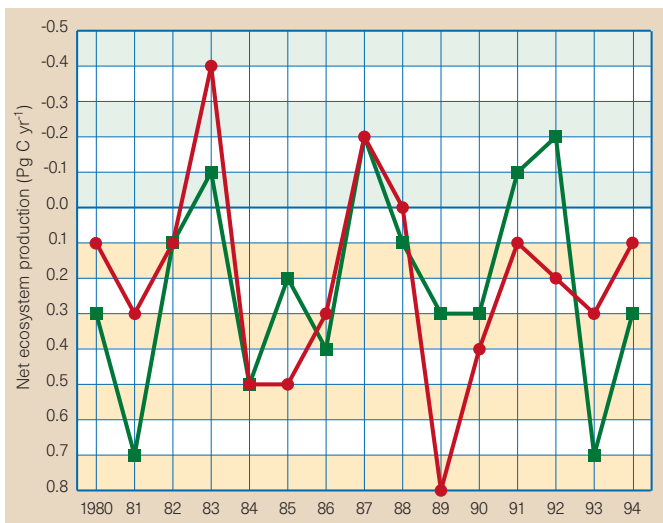


Figure 2 Two model calculations of the annual carbon balance of natural ecosystems in the Amazon Basin. The green trace shows the results of Tian *et al.*⁵; the red trace those from the Lund-Potsdam-Jena model (S. Sitch, B. Smith and J. Kaplan, unpublished results). Values below the zero line indicate periods when the region acted as a carbon sink.

however, in the wake of the United Nations climate conference in Kyoto, which took place this time last year and set targets for limiting emissions of greenhouse gases, locating sources and sinks for CO₂ has emerged as a policy issue with immense economic implications. Although the Kyoto protocol does not allow counting of 'natural' sinks, scientists have been quick to point out the illogicalities and perverse incentives that can be caused by incomplete carbon accounting¹¹.

Meanwhile, individual nations will not be slow to notice the implications of possessing such sinks, officially counted or not. There is therefore more justification than ever for devising accurate methods of assessing regional carbon balance. This is just one of the goals of the Brazilian-led international Large-scale Biosphere-Atmosphere Experiment in Amazonia, which will put the Amazon Basin in the biogeochemical spotlight for several years. In the starting phase of that project, the paper by Tian *et al.*⁵ serves as

a reminder, not only of the need for models for interpolation and prediction of carbon fluxes, but also of the paramount importance of more comprehensive measurements against which those models can be evaluated and refined.

I. Colin Prentice and Jon Lloyd are at the Max Planck Institute for Biogeochemistry, PO Box 100164, D-07701 Jena, Germany.
e-mails: colin.prentice@bgc-jena.mpg.de
jon.lloyd@bgc-jena.mpg.de

1. Heimann, M. *Ambio* **26**, 17–24 (1997).
2. Gifford, R. M. in *The Global Carbon Cycle* (ed. Heimann, M.) 159–199 (Springer, Berlin, 1993).
3. Lloyd, J. & Farquhar, G. D. *Funct. Ecol.* **10**, 4–32 (1996).
4. Grace, J. *et al.* *Science* **270**, 778–780 (1995).
5. Tian, H. *et al.* *Nature* **396**, 664–667 (1998).
6. Rayner, P. J. *et al.* *Tellus* (in the press).
7. Phillips, O. L. *et al.* *Science* **282**, 439–442 (1998).
8. Kindermann, J., Würth, G., Kohlmaier, G. H. & Badeck, F.-W. *Glob. Biogeochem. Cycles* **10**, 737–755 (1996).
9. Malhi, Y. *et al.* *J. Geophys. Res.* (in the press).
10. Braswell, B. H., Schimel, D. S., Linder, E. & Moore, B. III *Science* **278**, 870–872 (1998).
11. IGBP Terrestrial Carbon Working Group *Science* **280**, 1393–1394 (1998).

adult hermaphrodite (the wiring diagram) have already been described, and we can now add the genomic sequence. The numbers are impressive. The 97 Mb of DNA encode 19,099 predicted genes (three times as many as *S. cerevisiae*). Coding sequence accounts for 27% of the genome, with introns (an average of five per gene) accounting for another 26%. Although some gaps (containing less than 1% of the genome) remain to be closed, the sequence is sufficiently complete to allow meaningful insights into the worm genome. For example, the five autosomal chromosomes have a greater gene density in the centre than on the chromosomal arms (the X chromosome does not show regional differences). Matches to *C. elegans* complementary DNAs and to non-nematode proteins are also greatest in the central region, whereas tandem and inverted repeats are more common on the arms¹. These results support the suggestion that the chromosomal arms may be areas of rapid evolutionary change.

More insights come from the classes of genes in the *C. elegans* genome. Befitting a multicellular organism, the most frequently found motifs are for proteins involved in intercellular signalling and gene transcription¹. For example, there are 790 predicted seven-transmembrane receptors, 480 predicted zinc-finger proteins and 410 predicted proteins with protein-kinase domains. In one of the accompanying commentaries that provide examples of how the *C. elegans* data can be used^{5–9}, Chervitz *et al.*⁵ looked at orthologues — similar proteins with a presumed common ancestor — from the *C. elegans* and *S. cerevisiae* genomes. Most orthologues in the worm are needed for what these authors term core functions, such as intermediary metabolism, DNA-, RNA- and protein-metabolism, transport and secretion, and cytoskeletal structure. In contrast, yeast has no orthologues for many of the proteins involved in intercellular signalling and gene regulation in *C. elegans*.

For *C. elegans* researchers, the genome sequence and associated information (particularly Yuji Kohara's cDNA collection) have greatly accelerated our studies. Ten years ago, most genes for which mutations were known were cloned by tagging them with transposon insertions. Now, polymorphisms based on the genome sequence allow us to map genes rapidly and identify them by microinjection of rescuing DNA. Once identified, the sequence, genomic clones and, in many cases, cDNA clones are available for further study. Other homologues are available through database searches, and finding additional genes (or the lack thereof) through such searches — so-called *in silico* biology — is a wonderful perk of working with *C. elegans*.

The availability of sequence data has prompted many outside the field to investi-

Genome sequencing

The worm revealed

Martin Chalfie

In 1983, John Sulston and Alan Coulson began to construct a complete physical map of the genome of the nematode worm *Caenorhabditis elegans* (Fig. 1), and started what became known as the *C. elegans* Genome Project. At the time, several people wondered why John, who had just described all of the cell divisions in *C. elegans* (the cell lineage), was interested in this project rather than in a more 'biological' problem. He replied by joking that he had "a weakness for grandiose, meaningless projects". In 1989, as the physical map approached completion, the Genome Project, now including Bob Waterston and his group, embarked on the even more ambitious goal of obtaining the complete genomic sequence of the worm.

A summary of this work is now published

in *Science*¹, and the sequencing of the *C. elegans* genome has been declared "essentially complete". Far from being meaningless, this monumental achievement has yielded not only the largest genome sequenced to date — 97 mega base-pairs (Mb) compared with 12 Mb for the yeast *Saccharomyces cerevisiae*² — but, more importantly, the first genome sequence for a multicellular organism. The sequence^{3,4} provides the raw data from which to understand the biology of an entire animal, and with which to compare other genomes as they become known.

When Sydney Brenner pioneered the use of *C. elegans* for genetic studies, one attraction was the opportunity to describe aspects of its biology completely. The cell lineage and all connections of the 302 neurons in the

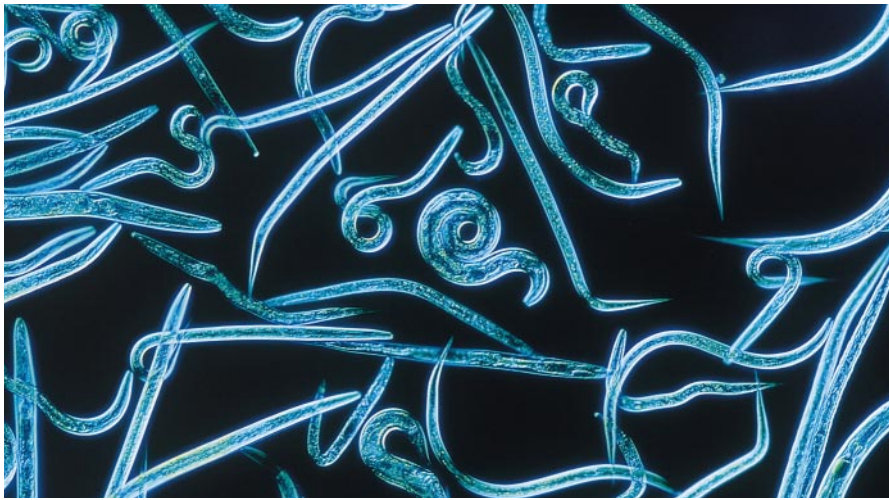


Figure 1 Worming out its secrets — the genome of the nematode *Caenorhabditis elegans* is now “essentially complete”.

gate *C. elegans* genes. Comparison with predicted human proteins shows that whereas 32% of worm proteins are similar to human sequences, 70% of human proteins identify similar sequences in *C. elegans*¹ — presumably this difference reflects the relative lack of information about human proteins. Those of us with established ‘worm’ labs are approached monthly, if not weekly, by people who have discovered that *C. elegans* contains a gene similar to one that they are studying. Indeed, although the worm lacks genes for some proteins, such as sodium channels, trk receptors and connexins, it has something for almost everyone. Genes have been identified for most known signalling proteins and transcription factors (see, for example, refs 6 and 7), as well as many genes that are similar to human disease-related genes.

But the benefits of the sequencing project go beyond merely identifying genes and working out the sequence. First, the *C. elegans* groups have shown that complex genomes can be sequenced. A strong argument can be made that the work of the *C. elegans* Genome Sequencing Consortium changed the course of the Human Genome Project, pushing it from a mapping mode to a sequencing mode. Second, the *C. elegans* project has been a model for how an efficient genome effort can be run. As far as I know, the consortium was the first to show the efficiency and cost-effectiveness of using an assembly-line approach rather than a researcher-orientated strategy. Third, the *C. elegans* Genome Project has stimulated the creation of powerful software needed to manipulate the genomic data. For example, the development of ACeDB (standing for ‘a *C. elegans* database’) by Richard Durbin and Jean Thierry-Mieg showed that infinite-record databases could be constructed. This database has been applied to other genome projects such as *S. cerevisiae* and the plant *Arabidopsis thaliana*. Another example is

GENEFINDER, constructed by Philip Green and LaDeana Hillier, which shows how coding sequences can be identified despite the complications of intron–exon structure and alternative splicing.

Fourth, and to me most important, the *C. elegans* Genome Project is a superb model of how sequencing can best serve the scientific community. When he first described the project, John Sulston remarked that one of his main goals was to promote an open and free exchange of materials. From the start, all data, clones and sequence were freely available, leading to the involvement of the entire *C. elegans* community. This cooperation accelerated the mapping phase of the

genome project by connecting the physical and genetic maps, and the openness was increased as the genome was sequenced. In a sense the *Science* paper, although a welcome benchmark, is somewhat anticlimactic because sequence data (even unfinished sequences) were made available as they were produced^{3,4}.

Obtaining the sequence — in itself an extraordinary achievement — sets the stage for the much larger project of analysing and understanding the genome. About 15% of the predicted genes have been confirmed by cDNA analysis. Confirmation, correction and annotation of the sequence will take many years, and will depend on our understanding of *C. elegans* biology. Even more important will be analysis of the regulatory pathways that link the genes, their products and their biological functions. Those studying *C. elegans* have always taken pride in the fact that they try to look at the whole animal. So, although publication of the sequence is an occasion for celebration and congratulation, it is also a call to get back to work. □

Martin Chalfie is in the Department of Biological Sciences, Columbia University, New York, New York 10027, USA.

e-mail: mc21@columbia.edu

1. The *C. elegans* Sequencing Consortium *Science* **282**, 2012–2018 (1998).
2. Goffeau, A. *et al.* *Science* **274**, 563–567 (1996).
3. http://genome.wustl.edu/gsc/C_elegans/
4. http://www.sanger.ac.uk/Projects/C_elegans/
5. Chervitz, S. A. *et al.* *Science* **282**, 2022–2028 (1998).
6. Bargmann, C. I. *Science* **282**, 2028–2033 (1998).
7. Ruvkun, G. & Hobert, O. *Science* **282**, 2033–2041 (1998).
8. Clarke, N. L. & Berg, J. M. *Science* **282**, 2018–2022 (1998).
9. Blaxter, M. *Science* **282**, 2041–2046 (1998).

Biomechanics

The quirks of jerks

M. A. R. Koehl

The back-and-forth motion of ocean waves can rip off plants and animals that are attached to the shore, and therefore affect the structure of bottom-dwelling communities in shallow coastal habitats^{1,2}. By analysing the forces on organisms in waves, we can try to understand why some body designs are resistant to being washed away, whereas others are more vulnerable. Flexible organisms such as kelp, for example (Fig. 1, page 623), can avoid large hydrodynamic forces by moving with the water as it sloshes shoreward and seaward. But, reporting in *Limnology and Oceanography*, Denny and colleagues³ point out the danger of going with the flow. A flexible organism that moves with the surrounding water gains momentum, which can impose an inertial force on the structure that attaches it when the tether yanks the organism to a halt. Through a series of simple mathematical models, the authors show that the inertial force on a flexible organism that has reached

the end of its rope can sometimes be larger than the hydrodynamic forces it experiences.

The mechanical flexibility of organisms attached to the substratum can reduce the fluid-dynamic forces that they must withstand by several mechanisms. Flexible organisms — such as trees in the wind and sea anemones in tidal currents — experience a considerable reduction in the drag force that pushes them downstream as they are reconfigured by the moving fluid into more streamlined shapes^{4,5}. Flexibility can also sometimes protect the attachment organs of bottom-dwelling creatures from bearing large forces if the creatures are subjected to the back-and-forth water motion of ocean waves. Wave-swept organisms are subjected to forces that depend on the instantaneous velocity (drag and lift) and acceleration (added mass force and virtual buoyancy) of the water relative to them^{4,6,7}. The sum of these forces on a rigid organism in waves varies with time as the water flows back and forth. In contrast, a



Figure 1 Flexible organisms like these kelp attached to wave-swept shores can avoid bearing hydrodynamic forces by moving with the water. But, according to Denny *et al.*³, they may experience large inertial forces if they are suddenly stopped short at the end of their leash.

flexible organism that can move with the water in waves can avoid being pulled by these hydrodynamic forces until it is strung out in the direction of flow and the water moves past it. The longer the organism relative to the distance the water moves before it flows back the other way, the more likely the organism is to avoid flow forces at times when accelerations and velocities are high⁸.

Denny *et al.*³ now point out that such a reduction in hydrodynamic force may not protect the structure that attaches the organism to the shore from potentially damaging loads if the inertial force on the organism is high when it reaches the end of its tether and suddenly stops moving with the flow. They used mathematical models to illustrate how the mass and size of an organism, the stiffness of its tether, and the period and peak velocity of the waves affect the inertial force on the organism when it reaches the end of its rope.

The authors propose that a dimensionless index, the 'jerk number', can be used to predict when inertial forces are important for a wave-swept organism. They used this approach to model the forces on kelp with upright, bending stipes, kelp with rope-like stipes, and mussels attached by stretchy, visco-elastic byssal threads. The jerk number is the ratio of the maximum inertial force that could act on an organism under particular wave conditions to the maximum hydrodynamic force that it would encounter if it were stationary and did not go with the flow. High jerk numbers (and, thus, large inertial forces) occur when the frequency of the back-and-forth motion of the waves is low relative to the resonant frequency of the organism. In such cases, flexibility is not a force-reducing mechanism. In contrast, if the period of the waves is short relative to the time it takes the organism to get strung out, the inertial force is low or zero because the mass never reaches the end of its tether. So,

the tether is not pulled and flexibility can reduce the forces experienced in waves.

The mechanical properties of the tissues that attach organisms to the substratum can have a big effect on their likelihood of experiencing large forces. For example, extensible kelp stipes act as shock absorbers, allowing the plants to withstand transient high forces⁹. The visco-elastic properties of an

organism's tissues cause different mechanical responses to environmental flows that vary on different timescales¹⁰. Denny and colleagues also explored tuning of an organism's material properties in time-varying flow environments. Their models revealed that the inertial loading of wave-swept organisms peaks at specific frequencies. So, the authors suggest that the structure and material properties of organisms might be altered — either by physiological response or during evolution — such that potentially damaging loads are avoided. An area for exploration might be how such tuning is maintained as these organisms grow and as flow conditions change with the seasons. □

M. A. R. Koehl is in the Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.

e-mail: cnidaria@socrates.berkeley.edu

1. Sousa, W. P. *Annu. Rev. Ecol. Syst.* **15**, 1353–1391 (1984).
2. Paine, R. T. & Levin, S. A. *Ecol. Monogr.* **51**, 145–178 (1981).
3. Denny, M. W., Gaylord, B., Helmuth, B. & Daniel, T. *Limnol. Oceanogr.* **43**, 955–968 (1998).
4. Koehl, M. A. R. *J. Exp. Biol.* **69**, 87–105 (1977).
5. Vogel, S. *Am. Zool.* **24**, 37–44 (1984).
6. Denny, M. W., Daniel, T. & Koehl, M. A. R. *Ecol. Monogr.* **55**, 69–102 (1985).
7. Denny, M. W. *Biology and Mechanics of the Wave-Swept Environment* (Princeton Univ. Press, 1988).
8. Koehl, M. A. R. in *On the Economy of Plant Form and Function* (ed. Givnish, T. J.) 603–634 (Cambridge Univ. Press, 1986).
9. Koehl, M. A. R. & Wainwright, S. A. *Limnol. Oceanogr.* **22**, 1067–1071 (1977).
10. Koehl, M. A. R. *J. Exp. Biol.* **69**, 239–268 (1977).

Structural biology

The ABC of a versatile engine

Michael J. Welsh, Andrew D. Robertson and Lynda S. Ostedgaard

Early in evolution an engine called a nucleotide-binding domain (NBD) was created. As evolution progressed, this engine proved quite adaptable — it now powers an amazing variety of machines. On page 703 of this issue, Hung and colleagues¹ describe the structure of the histidine permease NBD domain from *Salmonella typhimurium*, giving us a first look at this remarkable molecular engine. And, like any engine seen for the first time, we are eager to find out how it works.

Members of a large family of proteins called the ATP-binding cassette (ABC) trans-

porters² contain two NBDs together with two membrane-spanning domains (MSDs). These proteins are found in archaea, eubacteria and eukaryotes, and are increasingly recognized as the cause of human genetic diseases. The NBDs of different ABC transporters have considerable sequence similarity, and they all bind and hydrolyse ATP. In contrast, the transmembrane helices in the MSDs show little sequence similarity.

The organization of the NBDs and MSDs into a polypeptide chain varies between different ABC transporters. In many prokaryotic transporters, such as the *S. typhimurium*

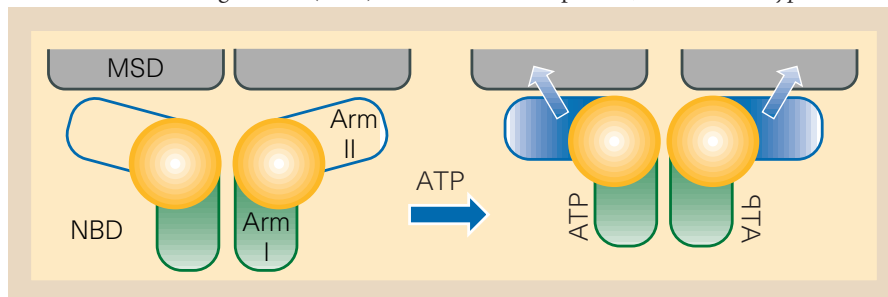


Figure 1 The nucleotide-binding domain (NBD) of *Salmonella typhimurium*. ATP-dependent conformational changes in the NBD are transduced into conformational changes in the membrane-spanning domains (MSD).

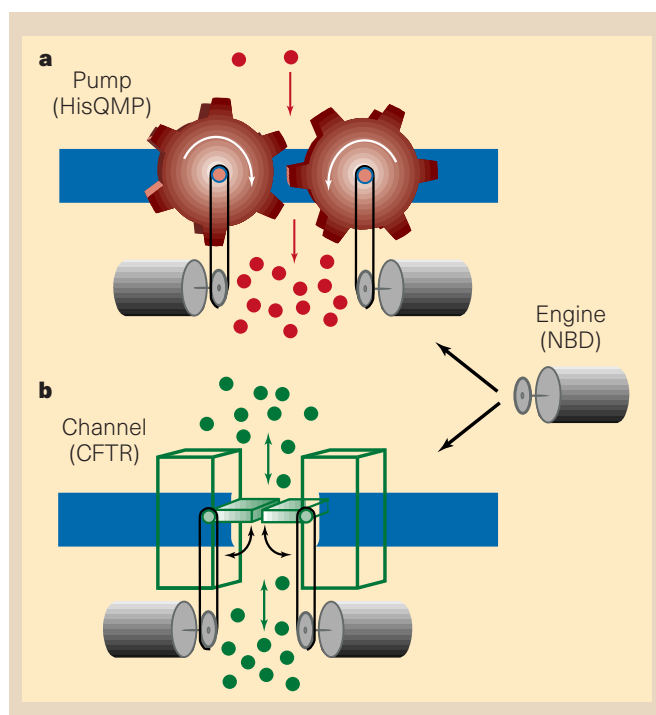


Figure 2 ABC transporter proteins. Members of this family contain two nucleotide-binding domains (NBDs) as well as two membrane-spanning domains. **a**, Some ABC transporters, such as the *Salmonella typhimurium* histidine permease, act as pumps. **b**, Others, such as the cystic fibrosis transmembrane conductance regulator (CFTR), act as channels. In reality, the pump and channel functions may not be so different, making ABC transporters interesting model systems in which to investigate this important relationship.

histidine permease³, each domain is encoded by a different gene, and individual domains are assembled into a membrane-bound complex. In many eukaryotic ABC transporters, however, a single protein includes both NBDs and both MSDs — the cystic fibrosis transmembrane conductance regulator (CFTR) is an example.

Just as a single type of engine can power a pump, a sewing machine or a door opener, so the NBD engines can power a variety of machines. ABC transporters act as pumps (histidine permease, for instance), as regulators of other membrane proteins (the sulphonyl receptor, SUR1) and as channels (CFTR). Thus, the structure reported by Hung *et al.*¹ is especially interesting because it probably applies to the NBDs of many ABC transporters.

Histidine permease is involved in a transport system that has been characterized by Ferro-Luzzi Ames and colleagues. The entire complex contains two MSDs (HisQ and HisM), two identical NBDs (HisP) and a periplasmic protein (HisJ), which binds histidine and presents it to HisQ and HisM for transport across the membrane. ATP hydrolysis by HisP powers the uptake and intracellular accumulation of histidine³. Hung and colleagues' crystal structure now shows that HisP has an unusual 'L' shape with two arms (Fig. 1). Arm I contains most of the residues that contact ATP, and arm II is an α -helical region that may interact with HisQ/HisM and the membrane. A six-stranded β -sheet sits at the angle of the L and contributes to both arms. This structure is consistent with earlier biochemical and genetic studies, supporting the idea that the structure reflects that of HisP in the membrane complex. Moreover, the structure provides the first

opportunity to ask questions about the molecular mechanics of the NBD engine.

How do HisP and other NBD engines convert the energy of ATP binding and hydrolysis into the conformational changes associated with membrane transport? The structures of HisP, guanine-nucleotide-binding (G) proteins and motor proteins such as myosin and kinesin suggest that they have similar mechanisms. All undergo conformational changes on binding and hydrolysis of nucleoside triphosphates, and all share two short regions of sequence similarity — the Walker A motif, or P (phosphate-binding) loop, and a Walker B motif containing a conserved aspartate residue^{4–6}. In HisP, these two motifs fall in structural regions remarkably similar to those in G proteins and motor proteins. Conformational switching in G proteins and motor proteins also involves a region that has an analogous counterpart in HisP.

Where is the movement in HisP? Standing on the sugar of the nucleoside triphosphates, looking at the γ -phosphate, one is pointed towards the region of conformational change in G proteins and motor proteins. In HisP the γ -phosphate points to arm II, suggesting that ATP binding and hydrolysis drive conformational changes in this arm. But how might movement in arm II couple to transmembrane transport? The MSDs determine the transport function of an ABC transporter, so perhaps arm II transduces conformational changes to the MSDs. This might involve the highly conserved LSGGQ motif at the end of arm II, which is partially buried by helix-6. The analogous helix in G proteins is linked to nucleotide-induced conformational changes^{5,6}. In HisP, ATP-induced conformational changes in helix-6 might expose the LSGGQ motif,

allowing interaction with the MSDs.

Why do ABC transporters contain two NBD engines? The structure shows us that HisP is a dimer (Fig. 1), and tells us at which face the monomers interact. Another ATPase that functions as a dimer is the Fe-protein of nitrogenase⁷. ATP-induced conformational changes re-orientate Fe-protein monomers with respect to one another, forming a new surface for protein-protein interaction and thereby amplifying small structural changes. The concept of structural amplification in dimers may also be applicable to NBDs. Alternatively, dimers might power sequential conformational changes. In CFTR, for instance, ATP binding/hydrolysis at one NBD may open the channel, whereas the interaction of ATP with the other NBD may close the channel.

Another question is whether Hung and colleagues' HisP structure can help us to understand how missense mutations in the NBD disrupt function in other ABC transporters (Fig. 2). Consider the CFTR chloride channel. Mutations associated with cystic fibrosis in arm I may disrupt channel gating by interfering with ATP binding or hydrolysis; thus, they may stop or slow the engines that control channel activity. Mutations in arm II (in the LSGGQ motif, for example) may inhibit channel activity by disrupting the interaction between the NBDs and MSDs; thus they may disengage the engines. Other cystic-fibrosis-associated mutations in either arm I or arm II can prevent normal folding and traffic to the cell surface — that is the case with the most common mutation, $\Delta F508$. It is interesting that second mutations in Arg 553 and Arg 555 of CFTR partially correct the $\Delta F508$ defect⁸, and they lie adjacent to the Phe 508 equivalent in the structure. Perhaps knowing the structure will provide the basis for pharmacological rescue of the processing defect.

What about the future? The structure of HisP, and the forthcoming structure of another NBD, RbsA (ref. 9), will allow us to model other NBDs. Moving beyond this, structural knowledge of intermediate states in the ATP hydrolytic cycle should tell us how the NBDs work. And, although it is asking a lot for a membrane protein, we would really like to know what the MSDs look like and how the NBDs drive them. □

Michael J. Welsh, Andrew D. Robertson and Lynda S. Ostedgaard are at the University of Iowa College of Medicine, Iowa City, Iowa 52242, USA.
e-mail: mjlwelsh@blue.weeg.uiowa.edu

1. Hung, L.-W. *et al.* *Nature* **396**, 703–707 (1998).
2. Hyde, S. *et al.* *Nature* **346**, 362–365 (1990).
3. Liu, P.-Q. & Ames, G. F.-L. *Proc. Natl Acad. Sci. USA* **95**, 3495–3500 (1998).
4. Walker, J. E., Saraste, M., Runswick, M. J. & Guy, N. J. *EMBO J.* **8**, 945–951 (1982).
5. Smith, C. A. & Rayment, I. *Biophys. J.* **70**, 1590–1602 (1996).
6. Vale, R. D. *J. Cell Biol.* **135**, 291–305 (1996).
7. Schindelin, H. *et al.* *Nature* **387**, 370–376 (1997).
8. Teem, T. L. *et al.* *Cell* **73**, 335–346 (1993).
9. Armstrong, S. R. *et al.* *ACA Meeting Abstr.* W0252 (1998).

Antigen presentation

A protease draws first blood

Matthew Bogyo and Hidde L. Ploegh

Biologists used to approach proteolysis with ambivalence and apprehension. Proteolytic cleavages cannot be predicted from a cloned sequence and, because it can destroy or confound one's favourite subjects, proteolysis is ignored at one's peril. During antigen processing, however, all of the proteolytic stops are pulled out because protein breakdown is necessary to provide molecules of the major histocompatibility complex (MHC) with their peptide cargo. On page 695 of this issue, Manoury and co-workers¹ add a new element to the story by showing that a protease, asparaginyl endopeptidase (AEP), is involved in MHC-restricted antigen presentation. This enzyme seems to deliver the initial blow to the tetanus toxin antigen, and leads to further processing of the antigen into epitopes that are recognized by the immune system.

For initiation of an immune response, antigenic fragments are usually presented to antigen-specific T lymphocytes by the products of the MHC. Most antigens that enter

the antigen-presenting cell via the endocytic pathway are presented by class II MHC products². Class II molecules are directed to the endocytic pathway by transient association with the invariant chain (Ii), which is the conceptual equivalent of a pro-piece — Ii must be removed for class II molecules to become fully active. This is done by lysosomal proteases, which destroy Ii and, at the same time, attack internalized antigens to convert them to suitably sized peptides. Removal of Ii is necessary not only for the class II MHC to acquire its peptide-binding properties, but also for its proper transport, as shown by experiments using pharmacological³ or even endogenous inhibitors of lysosomal proteases⁴.

Which lysosomal proteases can cleave the antigens and Ii? Attention was initially paid to the best-studied lysosomal cysteinyl and aspartyl proteases, such as cathepsins B and D. *In vitro*, at least, such enzymes can proteolytically remove Ii from the class II–Ii complex⁵, and generate T-cell epitopes from

intact protein antigens⁶. But, presumably, lysosomal proteases must work in concert to deliver a carefully orchestrated series of cuts if the correct peptide is to be generated. Antigen-presenting cells from knockout mice lacking cathepsins D, B or L show no obvious defects in antigen presentation^{7,8}, indicating that other proteases are involved in both removing Ii and converting the antigen into peptides.

One might assume, perhaps naively, that many lysosomal proteases are functionally redundant in view of their broad substrate specificities. Indeed — and consistent with such redundancy — deficiencies in lysosomal proteases are quite rare. Toulouse-Lautrec's short stature was attributed to a deficiency in cathepsin K, a minor lysosomal protease involved in bone remodelling⁹. But of the cathepsin knockouts studied so far, only animals lacking cathepsin D are sick, dying at around three weeks of age. Yet the fact that some lysosomal proteases have a restricted tissue distribution (such as cathepsin S, which is involved in proteolytic removal of Ii in professional antigen-presenting cells) raises the ante for the identification of further proteases in antigen presentation^{7,8}.

Consequently, the search has been on to implicate new proteases, many of them lurking in databases as cathepsin homologues. One of these has now been identified by Manoury *et al.*¹ in B cells as the cysteine protease AEP. A direct relative of the legumains, which were first discovered in plants, AEP is likely to be identical to the recently cloned legumain homologue from mammalian tissue¹⁰. The authors believe that AEP draws first blood and executes initial digestion of a protein antigen, tetanus toxin. It could help to unfold its target antigen to allow further digestion by the cathepsins.

To work out how a specific protease is involved in antigen processing, we need to be able to modulate its activity selectively. Inhibitors — often modified peptides that mimic an enzyme's substrates — transiently or permanently modify the active-site nucleophile, leading to a loss of activity. Manoury and co-workers used high concentrations of natural peptide substrates to block the activity of AEP *in vitro* and *in vivo*. But although the inhibitors were active against purified AEP, showing little or no effect on other cysteine proteases of the cathepsin family, caution is in order when transposing *in vitro* inhibitor studies to living cells. Many lysosomal cathepsins have similar substrate specificities *in vitro*, and inhibitors that are specific for single proteases are hard to find¹¹. So, small-molecule inhibitors that not only block activity, but also allow the molecular targets to be identified, are high on the wish-list. It is anyone's guess how many proteases with functional properties

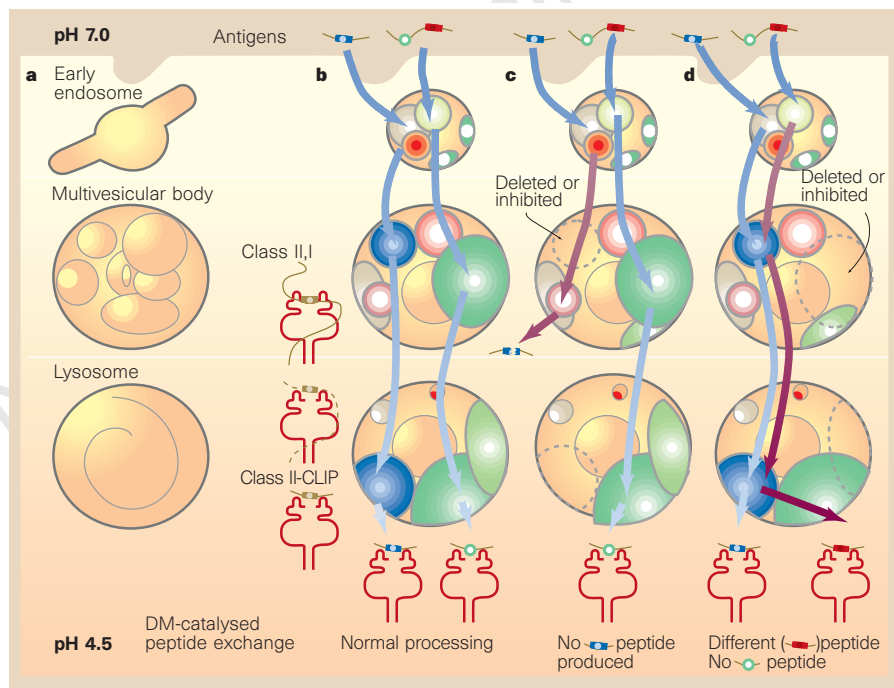


Figure 1 The endocytic pathway and proteolytic conversion of antigens. Coloured circles represent different specificities and activities of distinct endosomal/lysosomal proteases, and the geometric symbols in the antigen reflect different epitopes. **a**, The endocytic pathway. Newly synthesized class II major histocompatibility complex (MHC) molecules, complexed with the invariant chain (Ii), enter the endocytic pathway. Ii is removed by proteolysis to yield an Ii remnant, CLIP. The class II–CLIP complexes are substrates for peptide exchange catalysed by the class II-like product DM (ref. 2), and the result is formation of class II–peptide complexes. **b**, Under normal conditions (pale blue arrows), each antigen may sample distinct or (partially) overlapping proteases to generate T-cell epitopes. **c**, If a certain protease is ablated or inactivated, other proteolytic pathways may be accessed (burgundy arrows). Some epitopes may no longer be generated (for example, by cleavage within the epitope in the newly accessed pathway), whereas others are produced normally. **d**, Alternatively, enzyme ablation may produce epitopes that are not generated under normal circumstances. (For review see ref. 12.)

like those of AEP will be discovered. Nonetheless, the fact that antigen presentation is inhibited by competition for substrate at all, favours AEP as the protease that selectively makes the first cut in tetanus toxin.

An exciting prospect is the as-yet-hypothetical involvement of discrete, possibly unique, sets of proteases in converting certain self-proteins into the class II MHC-peptide complexes that trigger an autoimmune response. By identifying such proteases—and AEP could be amongst them—we could selectively inhibit them to blunt the severity of an autoimmune attack, if not prevent it altogether. Given the malleability of protease inhibitors in the medicinal chemist's hand, proteases may turn out to be the preferred targets for development of small-molecule immunomodulators. □

Matthew Bogyo and Hidde L. Ploegh are in the Department of Pathology, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA.

e-mail: ploegh@hms.harvard.edu

1. Manoury, B. *et al.* *Nature* **396**, 695–699 (1998).
2. Wolf, P. R. & Ploegh, H. L. *Annu. Rev. Cell Dev. Biol.* **11**, 267–306 (1995).
3. Neefjes, J. J. & Ploegh, H. L. *EMBO J.* **11**, 411–416 (1992).
4. Pierre, P. & Mellman, I. *Cell* **93**, 1135–1145 (1998).
5. Blum, J. S. & Cresswell, P. *Proc. Natl Acad. Sci. USA* **85**, 3975–3979 (1988).
6. van Noort, J. M. & Jacobs, M. J. *Eur. J. Immunol.* **24**, 2175–2180 (1994).
7. Villadangos, J. A., Riese, R. J., Peters, C., Chapman, H. A. & Ploegh, H. L. *J. Exp. Med.* **186**, 549–560 (1997).
8. Deussing, J. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 4516–4521 (1998).
9. Gelb, B. D., Shi, G. P., Chapman, H. A. & Desnick, R. J. *Science* **273**, 1236–1238 (1996).
10. Chen, J. *et al.* *J. Biol. Chem.* **272**, 8090–8098 (1997).
11. Chapman, H. A. *Curr. Opin. Immunol.* **10**, 93–102 (1998).
12. Geuze, H. J. *Immunol. Today* **19**, 282–287 (1998).

Superconductivity

An analogue of superfluid ^3He

Maurice Rice

Two letters to *Nature*^{1,2} published in the past few months, and one in this issue³ (page 658), report experiments that uniquely specify the unconventional superconductivity in strontium ruthenate (Sr_2RuO_4). They establish that it conforms to long-held views on how superconductivity driven by electron–electron (rather than electron–phonon) interactions should behave. These developments contrast with the heated debates that have swirled for a decade around the high-temperature copper oxide superconductors. Yet there are many similarities in the underlying electronic structures of the ruthenates and the copper oxides.

Shortly after the Bardeen, Cooper and Schrieffer (BCS) theory of conventional electron–phonon-driven superconductivity appeared in 1957, Kohn and Luttinger⁴ proposed a generalization to electron–electron interactions. The BCS theory starts from a normal metallic state with its ‘Fermi liquid’ of mobile electrons, with their interactions well described by the perturbative Landau theory. As the temperature is lowered an effective attraction between the electrons near the Fermi level arises from the exchange of phonons, or lattice vibrations, which grows in strength and overcomes the Coulomb repulsion. This net attraction binds two electrons to form an isotropic spin-singlet Cooper pair in a state of zero angular momentum (that is, *s*-wave). The origin of superconductivity is the condensation of the centre of mass motion of these pairs to form a coherent quantum state.

Kohn and Luttinger⁴ showed that even when the electron–electron interaction was predominantly repulsive it could still cause a subtle attraction, forming Cooper pairs in a

higher angular momentum state (such as *p*-wave or *d*-wave). This attraction was estimated to be very weak in simple metals. As a result, any such unconventional superconducting state would be difficult to observe due to a very low transition temperature (T_c), and then only in highly perfect crystals. But the attraction should be higher in transition metals and other materials, where the electronic states are concentrated close to the ions, thereby strengthening the interactions between electrons. In particular, if these interactions are strong enough to drive the metal to the verge of magnetic order, then the attractions would be further heightened by the exchange of magnetic fluctuations. The discovery in the 1970s of just such *p*-wave Cooper pairing in the neutral Fermi-liquid, ^3He , stimulated the search for an electronic counterpart.

In the 1980s, superconductivity was discovered in a class of rare earth and actinide compounds known as heavy fermion metals. Here the relevant electronic states (*4f* and *5f*) are even more compact and the proximity to magnetic order favours superconductivity (see ref. 5 for example). Surprisingly, the large class of transition metal oxides rarely shows superconductivity, although the variety of their spin and charge ordering attests to the presence of strong interactions.

Four years ago, Maeno and collaborators⁶ succeeded in preparing high-quality Sr_2RuO_4 and found a metallic state with hallmarks of a Landau–Fermi liquid, and a superconducting transition at $T_c \approx 0.7$ K. The crystal structure is that of the $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ copper oxide family. It consists of two-dimensional RuO_2 planes, and quantum coherence between the electrons in different planes occurs at $T < 50$ K. The most definitive



100 YEARS AGO

It is naturally difficult to obtain direct evidence as to how birds rid themselves of the indigestible parts of the fruit they eat. It is a question to which I have given some attention from its bearing on the dispersal of seeds. I have found large quantities of the seeds of hawthorn, dog-rose, mistletoe and ivy evidently voided by birds, as I incline to think generally as faeces, especially in the case of the hawthorn and ivy. Some large bird, I suppose the rook, consumes ivy berries largely in the spring, and gets rid of the seeds in what appears to be a mass of excrementitious matter. Many of these have not lost their vitality, and germinate readily in the same season. I have some thriving ivy plants obtained from such seed sown in 1896 and numerous seedlings this year of similar origin, the seed being sown on April 28, and coming up on June 7. I do not think much stress need be laid on the fact that much of the fruit swallowed is voided undigested, though the mistle-seeds I found were in a mass something like a lump of frog-spawn, with much of the pulp of the berry still adhering to each seed. I fancy birds and beasts, like many human beings, frequently swallow greedily far more than is good for them, especially when they light upon an abundant supply after enforced abstinence.

From *Nature* 15 December 1898.

50 YEARS AGO

In recent discussions on blinking, it has been taken for granted that, in a physical experiment, an object cannot be kept under continuous observation. The difficulty, however, is easily overcome. It is easy to acquire the habit of blinking the eyes alternately and, provided the period is made slightly shorter than that associated with normal blinking, no discomfort results. I have used this technique in the laboratory for more than twenty years, and have always assumed that it was generally known; but this, apparently, is not the case.

From *Nature* 18 December 1948.

Many more abstracts like these can be found in *A Bedside Nature: Genius and Eccentricity in Science, 1869–1953*, a 266-page book edited by Walter Gratzer. Contact Lisa O'Rourke. e-mail: l.orourke@nature.com

proof of the sharp Fermi surface in the normal state, which is central to the perturbative Landau theory, is the observation of periodic oscillations in the magnetization as the magnetic field is varied, known as de Haas–van Alphen oscillations. Mackenzie *et al.*⁷ succeeded in observing these oscillations and determined the shape of the Fermi surface, which agreed with band-structure calculations. The effective electron masses are unusually enhanced (factors of 3–5), indicating a strongly interacting Fermi liquid. The formal valence is Ru^{4+} , implying four electrons in bands of predominantly $4d$ character with $\{xy, yz, zx\}$ symmetry, all of which contribute to give three Fermi surface sheets with cylindrical topology.

The Ru^{4+} ion can be viewed as two holes in the $4d$ -complex and by Hund's rule their spins align parallel to give a triplet, $S = 1$, ion. Furthermore, the related cubic compound SrRuO_3 is a ferromagnetic metal. Such considerations led myself and Sigrist⁸ to speculate that the electrons in the Cooper pairs would have aligned spins leading to triplet superconductivity, and an odd orbital (for example, p -wave) wavefunction, analogous to ^3He and contrasting with the spin singlet d -wave pairing in the copper oxides.

Experiments determining the symmetry are now at hand. Luke *et al.*¹ performed spin-resonance experiments on muons implanted in Sr_2RuO_4 , which are a sensitive test for induced magnetic fields in the superconducting state. The observation of such fields

shows that the angular momentum of the Cooper pairs is not quenched by the crystal environment, and that the superconducting state has a broken time-reversal symmetry. Ishida *et al.*³ show that the Knight shift of the NMR peaks (which measures spin susceptibility) of oxygen nuclei remains unaffected by the transition to superconductivity. This is a clear sign of triplet pairing, because singlet pairing necessarily has a vanishing spin susceptibility as $T \rightarrow 0\text{ K}$.

There is another twist to the story. Specific-heat measurements⁹ showed that only about half of the Fermi surface had an energy gap in the superconducting state. A characteristic of Cooper pairs is that they can always scatter around a Fermi surface, conserving energy and momentum. As a result the whole Fermi surface should participate in electron pairing and possess an energy gap. Agterberg *et al.*¹⁰ suggested that the highly planar character of Sr_2RuO_4 would prevent scattering between $\{xy\}$ and $\{yz, zx\}$ sheets of the Fermi surface owing to their differing parities under reflection (that is, $z \rightarrow -z$). The question then is, which sheets are involved in the superconductivity?

Neutron scattering experiments by Rise-man *et al.*² provide the answer. Agterberg¹¹ calculated that, in a magnetic field, the vortex lattice of flux lines that penetrate the material should unusually have square symmetry, whose orientation depended on which Fermi surface sheets are involved in superconductivity. Rise-man *et al.* confirm the square symmetry, and from its orienta-

tion it follows that the pairing is occurring on the $\{xy\}$ -sheets. The results yield a complete description of the superconducting state — it is a two-dimensional analogue of the A-phase, one of the two distinct superfluid phases observed in ^3He .

So what have we learned? In a Landau–Fermi liquid, spin fluctuations can strengthen the pairing attraction, but it remains weak. Simultaneously, the mass enhancement lowers the characteristic energy scale of the attraction. As a result T_c is low. This combination of low T_c and unconventional superconductivity makes for extreme sensitivity to sample defects. Indeed, in the Sr_2RuO_4 -system a residual resistivity under $1\ \mu\Omega\text{cm}$ is required to observe superconductivity¹². Presumably many more unconventional superconductors could be found among the transition metal oxides if only they could be made perfect enough.

Although it is pleasing to see the original theory of superconductivity driven by electron–electron interactions verified, this route doesn't lead to high T_c . The copper oxides have a much higher T_c than the ruthenates; so what sets them apart? It cannot be the nearly two-dimensional character or the presence of strong spin fluctuations. These features are common to both. One difference lies in the formation in ruthenates of a well-defined Landau–Fermi liquid in the normal state ($T > T_c$). As emphasized by Anderson and others, such Landau character breaks down in the copper oxides. But this is a symptom rather than the cause of the

Atmospheric chemistry

Smelly cats

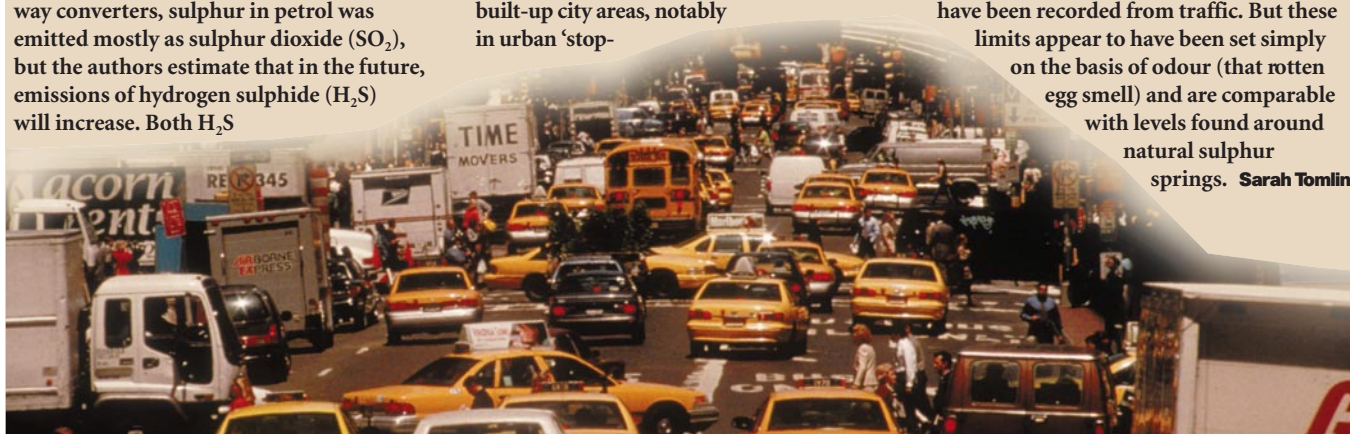
Since January 1993, all new cars in the European Union have had to be fitted with a three-way catalytic converter. Although three-way converters will lower emissions of certain harmful chemicals, such as oxides of nitrogen, S.F. Watts and C.N. Roberts raise the question of whether they may result in a shift in the balance of sulphur species, especially in cities (*Atmospheric Environment* 33, 169–170; 1999). Before the introduction of three-way converters, sulphur in petrol was emitted mostly as sulphur dioxide (SO_2), but the authors estimate that in the future, emissions of hydrogen sulphide (H_2S) will increase. Both H_2S

and SO_2 are respiratory toxins, but only levels of SO_2 are regularly monitored.

The data are both few and preliminary. Nonetheless, roadside measurements of H_2S in Britain and the United States seem to indicate that it reaches higher levels in Britain. Watts and Roberts speculate that this is due to different driving conditions, and point to previous studies that identified situations that favour higher emissions of H_2S instead of SO_2 . These include driving in built-up city areas, notably in urban 'stop-

and-go' traffic, and the first 10–20 km of a journey after cold starting, when the engine is rich in fuel. In Britain, as many as 40% of all car journeys are under 5 km and with increasing traffic congestion, the situation is likely to get worse.

Not only are emission data scarce, but it is hard to evaluate the levels of exposure that are harmful to humans. Concentration levels have been set for occupational hazards that are many times lower than have been recorded from traffic. But these limits appear to have been set simply on the basis of odour (that rotten egg smell) and are comparable with levels found around natural sulphur springs. Sarah Tomlin



difference. Perhaps the key distinction is the antiferromagnetic nature of the spin fluctuations in the copper oxides, whereas in Sr_2RuO_4 they are ferromagnetic. Only in the antiferromagnetic case do resonant valence bonds with their special stabilization of singlet pairs form. Maybe details of the dispersion of the energy bands also play a role.

We now have a complete description of the unconventional superconductivity of Sr_2RuO_4 . But it sharpens the most burning question in condensed matter physics today — just what is it that is so special about the high- T_c copper oxide superconductors? □

Maurice Rice is at the Institut für Theoretische Physik, ETH-Hönggerberg, CH-8093, Zürich, Switzerland.

e-mail: rice@itp.phys.ethz.ch

1. Luke, G. M. *et al.* *Nature* **394**, 558–561 (1998).
2. Riseman, T. M. *et al.* *Nature* **396**, 242–245 (1998).
3. Ishida, K. *et al.* *Nature* **396**, 658–660 (1998).
4. Kohn, W. & Luttinger, J. M. *Phys. Rev. Lett.* **15**, 524–527 (1965).
5. Mathur, N. D. *et al.* *Nature* **394**, 39–43 (1998).
6. Maeno, Y. *et al.* *Nature* **372**, 532–534 (1994).
7. Mackenzie, A. P. *et al.* *Phys. Rev. Lett.* **76**, 3786–3790 (1996).
8. Rice, T. M. & Sigrist, M. J. *Phys. C* **7**, L643–L698 (1997).
9. Maeno, Y. *et al.* *J. Low Temp. Phys.* **105**, 1577–1586 (1997).
10. Agterberg, D. F. *et al.* *Phys. Rev. Lett.* **78**, 3374–3378 (1997).
11. Agterberg, D. F. *Phys. Rev. Lett.* **80**, 5184–5188 (1998).
12. Mackenzie, A. P. *et al.* *Phys. Rev. Lett.* **80**, 161–165 (1998).

Apoptosis

Death deceiver

Douglas R. Green

The simple fact that cancer cells survive beyond their normal 'die by' date has only in the past decade been incorporated into our understanding of tumorigenesis. How transformed cells evade death provides insights not only into cancer, but also into the molecular mechanisms of apoptotic cell death, which eliminates those normal cells that might turn cancerous. On page 699 of this issue, Pitti *et al.*¹ report a newly discovered mechanism that tumours use to avoid a death signal — the production of a soluble molecule which binds to, and neutralizes, a trigger for apoptosis.

Fas (also known as CD95 or Apo-1) acts as a receptor on many different types of cell. When bound by its ligand (Fas ligand; FasL), Fas delivers a signal that leads to cell death (Fig. 1a). This is an important pathway for apoptotic cell death, although it is only one of several². Pitti *et al.* have now identified a molecule, which they call decoy receptor 3 (DcR3), that can also bind to FasL and block its binding to Fas. Moreover, DcR3 is genetically amplified in a number of lung and colon carcinomas, suggesting that Fas/FasL interactions limit cancer growth, and that cells expressing higher levels of DcR3 are more likely to become cancerous.

As well as amplified DcR3, transformed cells often produce a soluble form of Fas that can block the interaction between normal Fas and FasL³. Furthermore, Fas is mutated or downregulated in some cancers. These (and other) mechanisms interfere with the Fas pathway, which may then contribute to oncogenesis. In addition, although congenital defects in Fas expression or function cause a lymphoaccumulative syndrome, with accelerated autoimmune disorders, they are also associated with an increased risk of cancer in some individuals. These observations raise the intriguing possibility that Fas and FasL are tumour suppressors.

How, then, does FasL limit cancer such

that its inhibition by DcR3 might favour the disease? Pitti *et al.*¹ suggest that when DcR3 binds FasL, tumours may be able to evade surveillance by cytotoxic lymphocytes, which express FasL and can kill Fas-positive target cells (Fig. 1b). If so, such immune surveillance represents the selection for DcR3 amplification in tumour cells. Although this is certainly a possibility, cytotoxic T cells have another potent killer mechanism at

their disposal — they can release the contents of their cytotoxic granules, activating apoptosis downstream of Fas and other apoptotic pathways.

There are several other possibilities to explain how FasL restricts tumour growth. For instance, it is not only expressed on cytotoxic lymphocytes. Some tissues, such as the eye or testis, express FasL constitutively, and in other tissues (the intestinal epithelium, for example) FasL is inducibly expressed⁴. It is possible that tissue disruption caused by a tumour could trigger this expression and kill any Fas-sensitive transformed cells (Fig. 1c). Metastasis of lung melanoma has been shown to be suppressed by functional FasL (perhaps on the lung epithelium?) acting on tumour cells susceptible to Fas-mediated apoptosis⁵.

But FasL need not come from surrounding cells, as many tumours (including non-lymphoid tumours) themselves express functional FasL. DNA-damaging agents can trigger FasL production⁶, so increased expression of DcR3 might also be an adaptive response to cancer therapy (Fig. 1d). In a system designed to mimic the effects of chemotherapy on colon carcinoma cells, inhibition of FasL function preserved the ability of the cells to grow⁷. So increased expression of DcR3 may give tumour cells

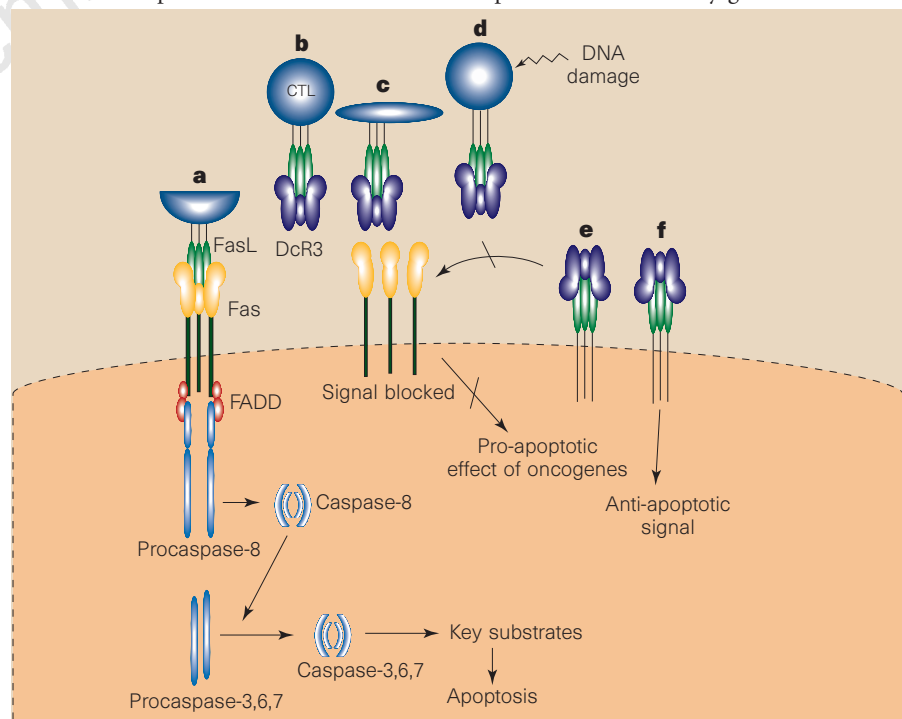


Figure 1 How DcR3 promotes tumorigenesis. a, Normally, trimerization of FasL-bound Fas recruits the cytoplasmic adaptor molecule FADD which, in turn, recruits procaspase-8, the precursor of an initiator caspase (cysteine protease with aspartic-acid specificity) involved in apoptosis. Two procaspase-8 molecules process one another, and assemble to form the mature, active caspase. This cleaves and activates other caspases (directly or indirectly²), which then orchestrate apoptotic cell death. b–f Anti-tumorigenic measures that may be blocked by the binding of DcR3 to FasL. b, Cytotoxic lymphocytes (CTLs) express FasL. c, Surrounding cells express FasL in response to tumour infiltration. d, DNA damage by therapeutic agents induces expression of FasL on tumour cells, which may kill neighbouring cells. e, Fas/FasL signalling normally contributes to the pro-apoptotic effects of oncogenes (such as c-Myc). f, Ligation of FasL delivers an anti-apoptotic signal to the cell.

some growth advantage during therapy.

Even more provocative is the idea that, by interfering with the Fas/FasL interaction, DcR3 directly contributes to oncogenesis. Expression of c-Myc induces apoptosis in fibroblasts when growth factors are limiting. Interfering with either FasL function or Fas signalling has been shown⁸ to prevent this pro-apoptotic effect of Myc, favouring Myc-driven proliferation (Fig. 1e). Moreover, in one transgenic model, expression of L-Myc in lymphocytes induced transformation only in mice lacking Fas⁹. It will be interesting to see whether DcR3 can cooperate with c-Myc, or other oncogene products, in transforming cells.

A final possibility is that DcR3 signals through FasL itself (Fig. 1f). Many tumours express FasL, which has a long cytosolic domain so that, paradoxically, it may act as a signalling receptor. Binding of Fas to FasL can, in some cases, induce a signal through FasL that results in cell proliferation¹⁰. If DcR3 can do this, it could promote tumour survival while countering the pro-apoptotic effects of Fas.

All of these possibilities are based on models that are either controversial or for which we have limited information. We do not know if any (or all) represent the real function of DcR3 amplification in promot-

ing tumorigenesis, or if they relate to its normal function. Fas and FasL belong to extended receptor/ligand families that include several proteins capable of triggering apoptosis — and several decoy receptors for these have been identified. Further, DcR3 binds FasL, despite having very low homology to Fas. Perhaps it also binds other ligands, or maybe other decoy receptors bind FasL and are selected in other tumours. Whatever the case, with each insight into how tumours manage to deceive death we get that much closer to learning who the relevant killers are. FasL seems to be a good candidate for one of them. □

Douglas R. Green is at the La Jolla Institute for Allergy and Immunology, 10355 Science Center Drive, San Diego, California 92121, USA.
e-mail: dgreen5240@aol.com

1. Pitti, R. M. *et al.* *Nature* **396**, 699–703 (1998).
2. Green, D. R. *Cell* **94**, 695–698 (1998).
3. Midis, G. P., Shen, Y. & Owen-Schaub, L. B. *Cancer Res.* **56**, 3870–3874 (1996).
4. Bonfoco, E. *et al.* *Immunity* **9**, 711–720 (1998).
5. Owen-Schaub, L. B., van Golen, K. L., Hill, L. L. & Price, J. E. *J. Exp. Med.* **188**, 1717–1723 (1998).
6. Friesen, C., Herr, I., Krammer, P. H. & Debatin, K. M. *Nature Med.* **2**, 574–577 (1996).
7. Houghton, J. A., Harwood, F. G. & Tillman, D. M. *Proc. Natl Acad. Sci. USA* **94**, 8144–8149 (1997).
8. Hueber, A. O. *et al.* *Science* **278**, 1305–1309 (1997).
9. Zornig, M., Grzeschiczek, A., Kowalski, M. B., Hartmann, K. U. & Moroy, T. *Oncogene* **10**, 2397–2401 (1995).
10. Suzuki, I. & Fink, P. J. *J. Exp. Med.* **187**, 123–128 (1998).

Immunology

The thymus in the age of retirement

Hans-Reimer Rodewald

The thymus — the central site at which T (thymus-derived) lymphocytes develop¹ — involutes with age^{2–6} (Fig. 1), so scientists have assumed that it is only functional early in life. On page 690 of this issue, however, Douek and colleagues⁷ report that the human thymus can generate new T cells throughout, and even late into, adult life. They also show that patients infected by human immunodeficiency virus (HIV) can generate new peripheral T cells after anti-viral therapy. These findings indicate that the thymus supports T-cell development throughout life, and that its function can be restored when it is relieved from disease-induced suppression.

In 1962, Miller¹ recognized that removing the thymus from newborn mice severely affected their immunological responses. Since then, immunologists have wondered whether or not the adult thymus still contributes to the peripheral T-cell pool by generating new T lymphocytes. This issue is of both conceptual and practical interest owing to the enormous size of the T-cell antigen receptor (TCR) repertoire — the sum of all TCR recognition specificities.

The TCR repertoire is established early in life and is estimated to contain at least around 10⁸ distinct specificities. Newly generated thymic emigrants colonize the

peripheral immune system as 'naive' T cells. After an immune response in which T cells provide help or cytotoxic activity, they either die or enter a memory compartment. Because T cells can be extremely long-lived, we cannot say whether the TCR repertoire in, say, a 70-year-old, was exclusively generated in his youth, or whether T-cells generated throughout life have contributed to the diversity of his TCR repertoire. Although naive and memory T cells can be distinguished by phenotype, there has been no assay for measuring thymic output in humans.

The basic mechanism for generating the TCR repertoire now provides the long-sought assay for thymic output. Two types of T cell are generated in the thymus. Most express a TCR composed of α and β chains, but a small population expresses a receptor composed of γ and δ chains. All of these chains are encoded by variable (V), diversity (D) and junctional (J) gene segments, which are rearranged during T-cell development in a process called V(D)J recombination⁸. This involves cleavage of DNA at the 'recombination signal sequences' that flank TCR gene segments in their germline configuration. The intervening stretches of DNA are excised, then the 'coding' ends are joined to form a functional TCR gene in the chromosomal DNA, and the 'signal' ends join to form extrachromosomal DNA circles termed TCR rearrangement excision circles (TRECs; Fig. 2)^{9,10}. The TCR- δ locus is embedded in the TCR- α locus, so TCR- δ sequences are specifically deleted in all $\alpha\beta$ T cells^{11,12}.

After successful TCR- α gene rearrangements, T cells expressing TCR- $\alpha\beta$ emigrate from the thymus without further cell division. In the life of a young T cell, a TCR- α gene rearrangement is a very recent event. Fortuitously, a molecular marker of this

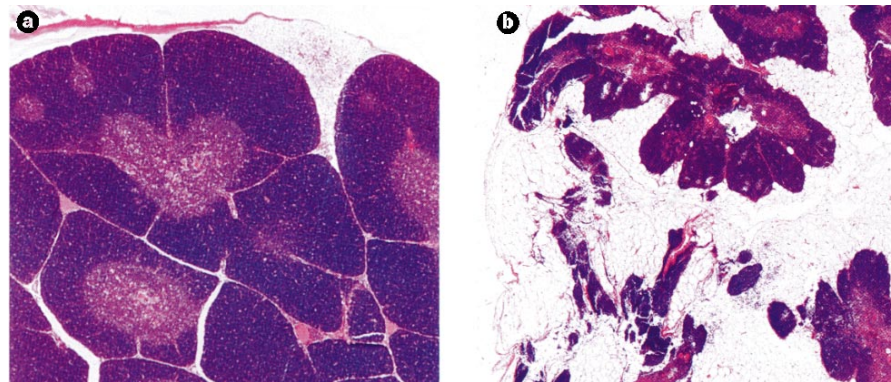


Figure 1 The size of the thymus decreases with increasing age. Until about 20 years of age, more than 80% of the thymus is composed of lymphoid tissue. With age, that proportion declines continuously until, from the age of about 40 onwards, only about 5% of the thymus is morphologically lymphoid⁶. Loss of lymphoid tissue — called thymic involution — seems to occur at the level of the thymic epithelium because, in mice, a young thymus can thrive when grafted into an old host, indicating that T-cell progenitors are active in aged bone marrow⁵. Histological examples of thymus sections from a child (a) and an adult (b) indicate dense thymic tissue in the young thymus only. The aged thymus is massively infiltrated by non-thymic tissue, mostly fat, but still harbours islands of lymphoid tissue, including both cortical and medullary 'units'. (Reproduced from ref. 15.)

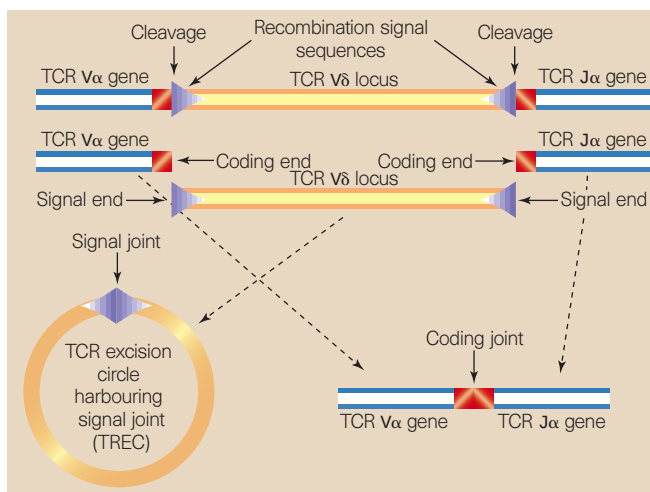


Figure 2 Representative V α -J α rearrangement at the T-cell receptor (TCR) $\alpha\beta$ locus. Deletion of germline DNA from the TCR- δ locus, followed by joining of the signal ends, results in a TCR excision circle (TREC). Douek *et al.*⁷ have analysed TRECs in human thymus and peripheral blood, and conclude that their presence in older people means that the thymus remains active.

event — a TCR- δ locus deletion circle — is carried from the thymus into the periphery. The TRECs are stable and cannot replicate. They are maintained in T cells after exit from the thymus, but are diluted out by cell division as the T cells get older¹³.

Douek *et al.*⁷ based their assay for thymic function on the detection of TRECs containing TCR- δ in T cells from peripheral blood. They detected TRECs in naive, but not memory, peripheral T cells expressing TCR- $\alpha\beta$ (but not TCR- $\gamma\delta$). The authors then quantified newly generated T cells (measured by levels of TRECs) as a function of age. They found many DNA circles at birth, but these declined roughly fivefold at age 35, and up to 50-fold at age 65. In contrast to common belief, *de novo* T-cell generation was detectable, even in people over the age of 70, albeit at low levels. So it appears that the T-cell pool is replenished by the thymus well into the 'retirement age'.

Remarkably, the relative numbers of TRECs detected by Douek *et al.* agree well with quantitative data on the lymphoid mass in the thymus with age⁶ (Fig. 1). Thus, in humans, the number of newly generated T cells that migrate from the thymus to the periphery seems to be proportional to the remaining thymus mass. The same conclusion has been drawn from experiments in the chicken, which has 14 thymic lobes (unlike humans or mice, which have two). By removing graded numbers of these lobes, a linear relationship was found between the mass and output of the thymus¹⁴. Douek *et al.* also measured TREC levels directly in DNA from the adult human thymus, because some T cells can be generated outside the thymus (in the gut and, perhaps, the liver). Moreover, peripheral T cells might maintain their TRECs for a long time if they do not divide. The results clearly show, however, that the number of active TCR rearrangements does not decline with age in thymus tissue.

Is this low-level generation of *de novo* T cells functionally significant? If we assume that, at birth, the thymus generates its full

repertoire of 10^8 specificities, an output of just 2% of this value would correspond to 2×10^6 specificities. Such an estimate ignores the rate at which T cells are produced in the aged thymus and the cellular turnover in the periphery. However, with the assay applied by Douek and colleagues for assessing thymic function in humans in a non-invasive way, we may now learn more about the TCR repertoire that is formed in the old thymus, the fidelity of negative and positive selection (failure of which may set the stage for more frequent autoimmune conditions with age), and the contribution of young T cells late in life to various T-cell subsets. It will also be important to understand how the function of the thymus is affected by treatments such as irradiation, steroids or anti-tumour therapy, some of which may cause permanent or temporary thymic involution. Finally, Douek *et al.* found that, in HIV-infected patients, highly active anti-retroviral therapy can lead to a recovery in the numbers of newly generated T cells in peripheral blood. This means that therapeutic restoration or maintenance of thymic function might be possible. □

Hans-Reimer Rodewald is at the Basel Institute for Immunology, Grenzacherstrasse 487, CH-4005 Basel, Switzerland.
e-mail: rodewald@bii.ch

1. Miller, J. F. A. P. *Proc. R. Soc. Lond. B* **156**, 415–428 (1962).
2. Hammar, J. A. *Arch. Anat. Physiol. Anat. Abt.* S91–S182 (1906).
3. Steinmann, G. *Curr. Top. Pathol.* **75**, 43–80 (1986).
4. George, A. J. & Ritter, M. A. *Immunol. Today* **17**, 267–272 (1996).
5. Mackall, C. L. & Gress, R. E. *Immunol. Rev.* **160**, 91–102 (1997).
6. Müller-Hermelink, H. K., Steinmann, G. & Stein, H. *Adv. Exp. Med. Biol.* **149**, 303–312 (1982).
7. Douek, D. C. *et al.* *Nature* **396**, 690–695 (1998).
8. Bogue, M. & Roth, D. B. *Curr. Opin. Immunol.* **8**, 175–180 (1996).
9. Fujimoto, S. & Yamagishi, H. *Nature* **327**, 242–243 (1987).
10. Okazaki, K., Davis, D. D. & Sakano, H. *Cell* **49**, 477–485 (1987).
11. de Villartay, J. P., Hockett, R. D., Coran, D., Korsmeyer, S. J. & Cohen, D. I. *Nature* **335**, 170–174 (1988).
12. Verschuren, M. C. *et al.* *J. Immunol.* **158**, 1208–1216 (1997).
13. Livak, F. & Schatz, D. G. *Mol. Cell. Biol.* **16**, 609–618 (1996).
14. Kong, F., Chen, C. H. & Cooper, M. D. *Immunity* **8**, 97–104 (1998).
15. Geneser, F. *Color Atlas of Histology* (Munksgaard, Copenhagen, 1985).

Daedalus

Boxing the compass

The Earth's magnetic field is maintained by 'homopolar dynamo' action in its molten interior. The metallic fluid thermally convects through the field; this induces an electric current which in turn sustains the field. The magnetohydrodynamic details are not exactly known, but the arrangement is obviously bistable. If field and current were reversed, the result would be an identical dynamo with opposite magnetic polarity. Indeed, the Earth's field has reversed repeatedly throughout its history. Its dynamo is only just stable. It may even be a very low-frequency alternator.

In the late Triassic period, the Earth suffered a set of meteoric impacts, close together in time. Two of them have opposite magnetic signatures (*Nature* 395, 126; 1998). It seems unlikely that two independent impacts (if such they were) should happen to straddle a magnetic reversal. Unless, says Daedalus, the first impact *caused* the reversal.

A bistable oscillator, such as a multivibrator, set up in either of its states, becomes steadily more and more sensitive to perturbation. Ultimately an arbitrarily small disturbance will unbalance it, when it will 'flip' into its other state. If the Earth's dynamo were approaching its sensitive phase, a powerful meteoric impact could well trigger its reversal. There is a claim that a new geomagnetic reversal is now approaching. So Daedalus proposes to trigger it pre-emptively.

A meteoric impact, he points out, is a very blunt instrument. Very little of its energy can plumb the geomagnetic depths. A pattern of underground nuclear explosions, cunningly staggered to create a big descending and converging shock-wave, could be far more effective. The bombs should be sited and aimed to tickle the dynamo in its most sensitive region. The resulting magnetic tremor would be highly informative, and might just start a magnetic reversal.

Even with the best modern geophysical understanding, the experiment would be a gamble. If it worked, its outcome would be an even bigger gamble. But the approaching crises of global warming and ozone-layer destruction demand bold action. Past magnetic reversals have been correlated with climate cooling and falls in sea level; maybe the trick will work again. The disturbance of the Earth's field would also change the impact of the Solar wind on the ionosphere, and with luck might repair the damaged ozone layer.

David Jones

Obituary

Nicholas Kurti (1908–98)

Low-temperature physicist and 'gastrophysicist'

One of the distinctions that pleased Nicholas Kurti most was appearing in the *Guinness Book of Records* as the creator of the then world-record low temperature, about a hundred-thousandth of a degree above absolute zero. That combination of scientific skill with a sense of fun epitomized the man. The work on low-temperature physics was conducted in the Clarendon Laboratory in Oxford, then headed by F. E. Simon with whom Kurti had a long connection. The lowest possible temperatures were reached in the early 1950s using the technique of adiabatic demagnetization to align the spins of atoms. To achieve this demanded enormous amounts of power, and Kurti's characteristic drive and determination enabled him to obtain for his experiments the generator that had powered the city of Manchester's tramway system.

The first 18 years of his life were spent in his native Budapest, under four totally different systems of government: a constitutional monarchy during the reigns of Franz Josef and Charles IV; a liberal republic in 1918 for five months, followed by a communist Soviet republic; and then for the next seven years a pseudo-democratic, near-fascist kingdom, ruled by a regent who was also an admiral.

It was anti-semitism that led him, along with many members of the Jewish middle-class, to seek his future outside Hungary. Unlike later emigrations this was not an escape from terror, but rather the talented leaving for better opportunities. In Kurti's case he went to Paris to study physics and chemistry, spending two years at the Sorbonne, while living in a hotel in the Place du Panthéon, which left him with an abiding love of France.

The defining period of his scientific life was the postgraduate time he spent in Berlin when he was part of a brilliant period of twentieth-century physics. He attended the Wednesday colloquia in the department of physics at the University of Berlin organized by Max von Laue, when the front row was occupied by Einstein, Hertz, Nernst, Planck and Schrödinger (the 'big guns' in Kurti's terminology), with the 'small fry' present including Ladenburg, Pringsheim, Simon, Szilard, London and Wigner.

He did his thesis work under F. E. (later Sir Francis) Simon in the department of physical chemistry in Berlin, completing

his thesis in two years. This was despite destroying his laboratory in an explosion when using a magnet cooled in liquid hydrogen. He was examined by Nernst and Schrödinger, in whose home the examination took place; this event was remembered by Kurti as a comfortable, shirt-sleeved, hour-long, interesting and instructive chat. Schrödinger gave him a "*sehr gut*" (very good) and Nernst a "*vorzüglich*" (excellent) resulting in a *magna cum laude* doctorate. Apparently the highest award of *summa cum laude* tended to be reserved for students who were poor or who were university teachers, because the university reimbursed the fees of those with *summa cum laude* and these two categories of students did not pay fees.

Kurti and Simon continued to work together from 1931 to 1933 at the Technische Hochschule in Breslau, where their Jewish origins became a problem with Hitler's rise to power. Fortunately the intervention of Lindemann (later Lord Cherwell) brought both of them to Oxford where the influx of Jewish physicists had an enormous and positive effect on the Clarendon Laboratory. The two were an ideal combination, Kurti's experimental skills complementing Simon's deep understanding of thermodynamics. During the Second World War they again collaborated, on the atomic bomb project and in particular on isotope separation.

Kurti became a fellow of Brasenose College at Oxford in 1947, professor of physics in 1967, a fellow of the Royal Society in 1956 (vice-president in 1965) and was appointed Commander of the British Empire in 1973. He sat on numerous commissions and had a highly successful scientific career, retiring formally in 1975. But that only gave him wider opportunities to demonstrate his vigour and relentless pursuit of a range of activities. Among other campaigns he battled with British Rail to reorganize the parking at Oxford station, challenging them to prosecute him for breaking the barrier that trapped motorists in the evening until operated by a railway official.

He had a serious interest in scientific history and was instrumental in setting up the archive of scientists' papers with the Royal Society and Royal Commission on Historical Manuscripts. He was also among those who pressed successfully

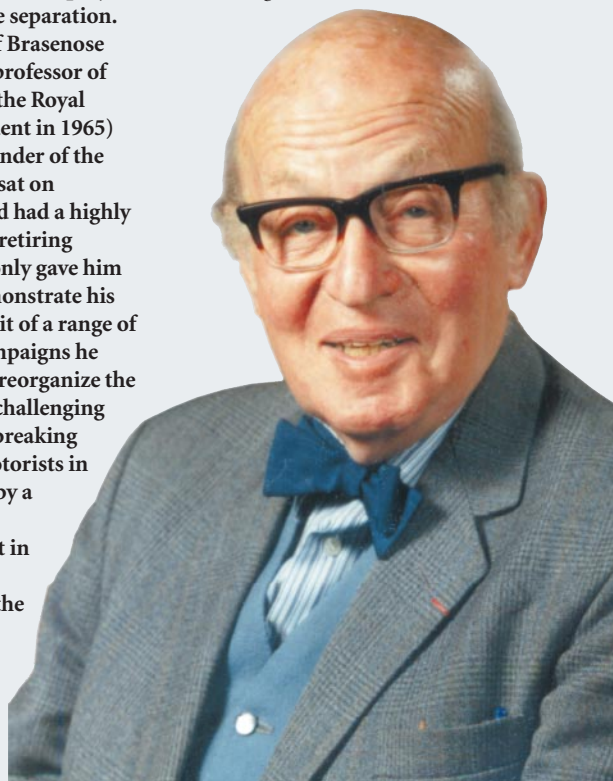
for the release of the recorded conversations of the German physicists interned at Farm Hall in Britain immediately after the end of the war, including their reactions to the Hiroshima bomb.

He was a highly talented musician with a love of opera, but chief among his pleasures was cooking, although very much influenced by his background as an experimental physicist. He liked to call his new topic 'gastrophysics' and presented a spectacularly popular lecture entitled "The physicist in the kitchen" at the Royal Institution and on television. One of his creations was an ice cream transparent to microwaves, enabling him to achieve an 'inverted baked Alaska', frozen on the outside but with hot jam in the centre.

Up until after his 90th birthday he cycled to his laboratory office early each day where among other things he wrote letters to newspapers, notably *The Times*, on a wide range of topics. Although he married quite late in life, he and his supportive wife Giana celebrated a golden wedding anniversary and together edited the Royal Society cookbook entitled *But the Crackling is Superb* (Adam Hilger, 1988). Nicholas Kurti, who died on 24 November, was not just a link with the heroic era of physics but a true original and creator of fun.

W. Graham Richards

W. Graham Richards is in the Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QT, UK.
e-mail: graham.richards@chem.ox.ac.uk



GODFREY ARGENT

Stilled splashes

The physicist Arthur Worthington was intrigued by the beauty to be found in photographs of splashes produced when bodies of various shapes and sizes fall into fluids. The legacy of his enthusiasm is with us today.

Martin Kemp

Some of the images made visible by modern science are at once so arresting and so beautiful that they have come to be regarded as supremely representative of natural forms and processes. The momentary coronet of the splash made by a spherical object plunging into a pool of milk has entered general consciousness to such effect that it serves as the logo of Milk Marque, the British company that replaced the Milk Marketing Board in 1994 and which is responsible for distributing more than 6 billion litres of milk per year.

The unexpected and wonderfully complex configurations of 'simple' splashes were first revealed in 1908 by Arthur Worthington, professor of physics at the Royal Naval Engineering College in Devonport, where research into the behaviour of water and the dynamics of projectiles was of obvious relevance.

Worthington's *A Study of Splashes* and his photographic montages, now in the National Museum of Photography, Film and Television in Bradford, are under-recognized classics of scientific photography. Each fleeting phase in splashes formed by falling bodies of various kinds and sizes was effectively stilled

in the light of a spark lasting less than three millionths of a second. Sets of splashes

formed under identical conditions were successively photographed by a plate camera with uncovered lens in a darkroom, with each successive splash captured at an interval of around one hundredth of a second later than its predecessor — on the assumption that each splash passes through essentially the same sequence of phases.

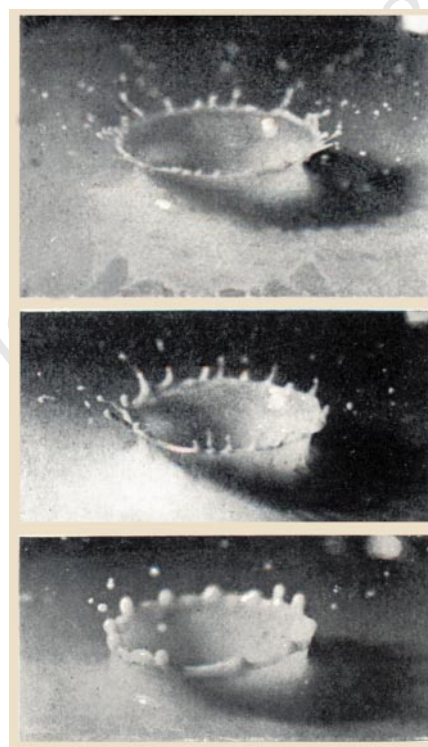
Worthington was justifiably excited by the "exquisite forms the camera has revealed". On one hand, the beguiling sequence of fluid configurations appears "orderly and inevitable", while on the other, "it taxes the highest mathematical powers to elucidate", requiring a precise knowledge of "the real path of any particle".

The sensitivity of the process to its initial conditions was confirmed when smooth and rough spheres of identical dimensions and weight were shown to generate quite different shapes, one more crater-like and the other more roundly protuberant. We now understand that the necessary 'mathematical powers' to model such dynamic systems reside in computer-driven chaos theory rather than with the analytical tools available at the turn of the century.

Although Worthington's experiments were drawn into a wider arena by their illustration in D'Arcy Thompson's *On Growth and Form* in 1917, in which the Scottish biologist characteristically intuits morphological analogies with medusoids and hydroid polyps, Worthington's pioneering work was later submerged by the entrepreneurial flair of Harold Edgerton at MIT's 'Strobe Alley' during the 1930s.

Edgerton so perfected his stroboscopic system that it could deliver 3,000 images per second. Alert

A Milk Marque tanker showing the familiar 'splash of milk' logo.



Arthur Worthington's "The splash of a sphere" from *A Study of Splashes* (Longmans, Green, 1908).

to the commercial potential of such striking photographs as that of a bullet passing through an apple, Edgerton vigorously promoted his pictures to the public, not least through his 1939 book *Flash* and his film *Quicker 'n a Wink*. It was one of Edgerton's spectacularly well-defined prints of the splash coronet that was accorded pride of place in Thompson's 1948 revision of his classic text, and, in its luridly coloured 1957 version, that entered the collections of museums, adorned posters and sold postcards.

In selecting a tidied-up graphic version of the milk coronet, seen by implication from below rather than looking down on the surface, Milk Marque is clearly relying upon our ability to see both the regal crown and the photographic splash, exploiting our receptiveness to visual puns. Seeing the splash makes more demands on us than seeing the stock icon of the crown, and it transpires that, famous though the Worthington-Edgerton image has become, not everyone picks up the ingenious allusion to the stilled splash in the logo that is familiar on the roads of Britain. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.



Do all freshwater eels migrate?

The catadromous migration of freshwater eels, in which they migrate from freshwater streams to the sea to spawn¹, is widely accepted. The proportion of time spent in freshwater and ocean habitats can be determined by studying the ratio of strontium (Sr) and calcium (Ca) in the otoliths (earstones) of the eels. Here we use this technique to show that Atlantic and Pacific eels collected in the ocean have spent their entire lifetime there and have never migrated into fresh water. This finding indicates that freshwater eels need not be catadromous, and that populations from the sea contribute primarily to future recruitment.

The Atlantic eels *Anguilla anguilla* (European eel) and *A. rostrata* (American eel) spawn in the Sargasso Sea and their transparent, leaf-like larvae, called leptocephali, move by ocean currents up the coast of Europe and North America, respectively. After metamorphosing into 'glass' eels, they are believed to migrate up freshwater streams, where they grow to the adult 'silver' stage, begin maturing their gonads, and then migrate back downstream, through the ocean to the Sargasso Sea, where they spawn and die. The Japanese eel, *A. japonica*, is thought to adopt a similar life

cycle, with adults spawning near the Mariana Islands², from where the larvae migrate to the Asian coast to enter freshwater rivers.

Despite considerable research, major components of the biology of these eels remain unknown. For instance, what factors determine the entry of glass eels into fresh water? Do all freshwater eels migrate into fresh water and, if so, why are many immature yellow-stage eels found in the ocean (more than 80 per cent of eels in the commercial ocean catch of the North Sea are yellow)? The general assumption that these yellow eels are washed from rivers by floods has not been tested.

We collected 49 specimens from both freshwater (9 specimens from the River Elbe in Germany and 10 from the River Tone in Japan) and ocean habitats (18 from the North Sea and 12 from the East China Sea) for both Atlantic and Pacific species.

Analysis of the Sr/Ca ratio in the otolith provides information about the time spent in each habitat^{3–7}, as fresh water has little Sr, whereas ocean has a high concentration⁸. We used synchrotron X-ray analysis⁹ to measure the Sr/Ca ratios in individual eel otoliths. Two-dimensional images showed a remarkable difference in Sr/Ca ratio

between ocean and river samples (Fig. 1a,b). We used these data to construct a life-history transect of individual ontogeny (Fig. 1c).

All river samples contained a region of high Sr 200–300 μm in diameter in the core region of otoliths. This corresponds to the leptocephalus stage of migration from the spawning ground to the juvenile rearing habitat. These central cores were surrounded by an area with a much lower Sr/Ca ratio, which corresponds to metamorphosis¹⁰ of the juvenile eels and entry into a freshwater habitat. The low Sr value continued until the eels were caught.

In contrast, all specimens collected in the North Sea and the East China Sea had been exposed to much higher amounts of Sr throughout their life. They showed no evidence of having migrated into freshwater habitats, although the high Sr/Ca ratio recorded during the leptocephalus stage decreased after metamorphosis, as in the freshwater samples. There was no obvious difference between the maturing silver eels (7 specimens) and the actively foraging yellow eels (11 specimens), both of which have an exclusively marine life history.

None of the 30 specimens collected in an ocean habitat, regardless of species, sex or developmental stage, had any freshwater history, which demands a new interpretation of the life history of eels. It is widely accepted that eels are catadromous and have a freshwater growth stage, but this is not an obligate migratory pathway; they should instead be seen to have a facultative catadromy with ocean residents as an ecophenotype.

Furthermore, because none of the 19 maturing silver eels from the ocean derived from freshwater migrants, there seems to be relatively little recruitment from freshwater habitats back to the ocean and the spawning grounds. Our finding that all 11 yellow individuals were fully marine, and that more than 80 per cent of all eels in the North Sea are yellow, indicates that eels with a primarily ocean-based life history are contributing to future recruitment, and that genes are maintained between generations only by eels that grow in the sea.

Katsumi Tsukamoto*, Izumi Nakai†, W.-V. Tesch‡

*Ocean Research Institute,
University of Tokyo,
Nakano-ku, Tokyo 164-8639, Japan
e-mail: ktpc@ori.u-tokyo.ac.jp

†Department of Applied Chemistry,
Science University of Tokyo,

Kagurazaka, Shinjuku, Tokyo 162, Japan

‡Biologische Anstalt Helgoland,
Notkestrasse 31, D-22607 Hamburg, Germany

1. McDowall, R. M. *Diadromy in Fishes* (Croom Helm, London, 1988).
2. Tsukamoto, K. *Nature* 356, 789–791 (1992).

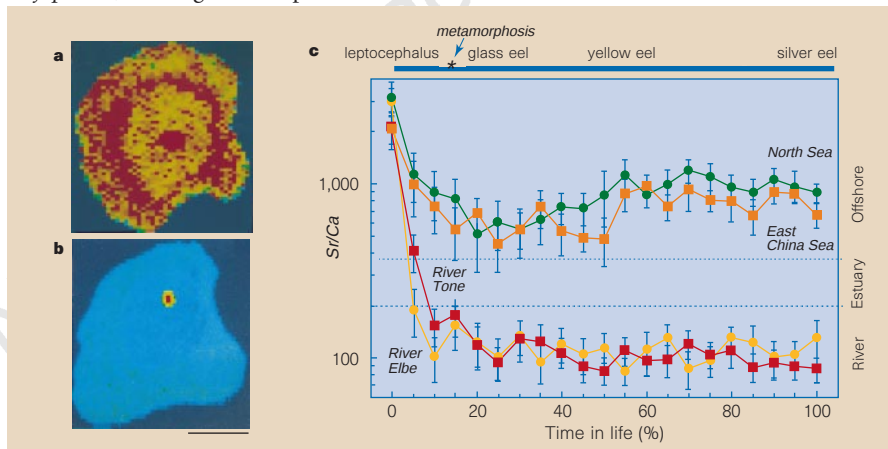


Figure 1 Two-dimensional images of silver eel otolith and the patterns of Sr/Ca ratio in freshwater eel otoliths. **a**, Ocean sample from the North Sea; **b**, freshwater sample from the River Elbe. Red indicates high Sr/Ca intensity. Scale bar, 1 mm. **c**, Filled circles, European eel captured in the North Sea in November 1988 (total length, 40.8–61.2 cm; $n=18$ (11 females, 7 males)); open circles, European eel collected in the River Elbe in November 1990 (total length, 40.1–78.0 cm, $n=9$ (7 females, 2 males)); filled squares, Japanese eel collected in the East China Sea in November 1994 (total length, 42.4–81.5 cm; $n=12$ (8 females, 4 males)); open squares, Japanese eel collected in the River Tone in November 1992 (total length, 61.2–105.0 cm; $n=10$ (8 females, 2 males)). The North Sea sample includes 11 yellow-stage specimens; other samples are all silver. The Sr/Ca value at each 5% interval was interpolated and the mean and standard deviation obtained. Sagittal otoliths of each fish were embedded in polyester resin and prepared for measurements of Sr and Ca content using synchrotron radiation-induced X-ray fluorescence (XRF) analysis by grinding and polishing the sample from both sides to the midplane of 30 μm thickness with polishing papers⁹. XRF measurements were made at beamline 4A of the Photon Factory, KEK, Tsukuba, Japan, using an energy-dispersive XRF system with a Si (Li) detector. Samples were excited by 17-keV X-rays obtained by a silicon double crystal monochromator. A fine parallel beam of 120 μm width and 100 μm height was obtained using vertical and horizontal slits. Two-dimensional analysis of a sample was made in air by step scanning the sample on a pulse-motor-controlled x-y stage, and X-ray fluorescence intensities of Sr K α and Ca K α lines were measured at each point; measuring time, 3 s per point; step size, 50 μm in x and y directions.

3. Casselman, J. M. in *Proc. 1980 North Am. Eel Conf.* (ed. Loftus, K. H.) 74–82 (Ontario Ministry of Natural Resources, Ontario, 1982).
4. Radtke, R. L. *Comp. Biochem. Physiol. A* **92**, 189–193 (1989).
5. Kalish, J. M. *J. Exp. Mar. Biol. Ecol.* **132**, 151–178 (1989).
6. Secor, D. H. *Fish. Bull. US* **90**, 798–806 (1992).
7. Tzeng, W. N., Severin, K. P. & Wickström, H. *Mar. Ecol. Prog. Ser.* **149**, 73–81 (1997).
8. Angino, E. E., Billings, G. K. & Anderson, N. *Chem. Geol.* **1**, 145–153 (1966).
9. Nakai, I., Iwata, R. & Tsukamoto, K. *Spectrochim. Acta B* (in the press).
10. Otake, T., Ishii, T., Nakahara, M. & Nakamura, R. *Mar. Ecol. Prog. Ser.* **112**, 189–193 (1994).

Trick or treat from food endocannabinoids?

The discovery of the endogenous cannabinoid *N*-arachidonylethanolamine (anandamide)¹ and other *N*-acylethanolamines (NAEs) in chocolate² has led to speculation that the purported rewarding properties of cocoa are due to the presence of compounds “that could act as cannabinoid mimics”². This observation raises some important questions. First, are NAEs and anandamide, or the ‘endocannabinoid’ 2-arachidonoylglycerol (2-AG)³, present in widely consumed foods (such as milk) that are less ‘rewarding’ than chocolate? And second, to what extent do these compounds reach the bloodstream and exert pharmacological effects when consumed orally? We believe that the content of endocannabinoids in foods, and in cocoa in particular, is not sufficient to produce cannabis-like effects in mammals.

We purified NAEs and 2-AG from various foods (including mature human, bovine and goat milk, and cocoa at various stages of processing) and quantified them by gas chromatography and mass spectrometry. Milk was the only food analysed in which an endocannabinoid, 2-AG, was found in relatively high concentrations ($0.33 \pm 0.11 \mu\text{g ml}^{-1}$ in human milk, for example). The level of NAEs varied between 0.003 and $0.024 \mu\text{g ml}^{-1}$, with anandamide being the least abundant compound. Oleamide, a sleep-inducing substance⁴ that inhibits the hydrolysis of anandamide^{5,6}, was also present (0.055 – $1.27 \mu\text{g ml}^{-1}$).

In cocoa-derived samples, we detected

NAEs (0.01 – $5.8 \mu\text{g per g}$) and oleamide (0.17 – $6.0 \mu\text{g per g}$), but no or very little anandamide and no 2-AG. NAE levels are much lower in unfermented cocoa beans than in cocoa powder (which contained less than $0.003 \mu\text{g per g}$ anandamide). Tiny amounts of anandamide in cocoa could therefore be explained as artefacts of processing². Like all higher plants, cocoa plants cannot synthesize arachidonic acid or its derivatives⁷. Notably, in their NAE and oleamide content, cocoa and dark chocolate are similar to other plant-derived foods (including soybean, hazelnuts, oatmeal and millet), in which we detected up to 2.3 , 1.1 and $2.8 \mu\text{g per g}$ oleamide, *N*-oleoyl-ethanolamine and *N*-linoleoyl-ethanolamine, respectively.

To establish whether endocannabinoids survive their passage through the digestive system, we assayed anandamide and 2-AG after oral administration in a series of *in vivo* tests that are used widely to assess cannabinomimetic activity⁸ (Table 1). The two compounds were active in four of the five behavioural tests, but only at very high concentrations relative to those in foods. In contrast, a smaller dose of the psychoactive component of marijuana, $(-)\Delta^9$ -tetrahydrocannabinol (Δ^9 -THC), was sufficient to give a strong effect in all tests.

Because intraperitoneal anandamide and 2-AG are active in the same tests at doses 20- to 60-fold lower^{3,9}, these results indicate that only 1.6–5% of the orally administered compounds enter the bloodstream, probably owing to the high levels in the gastrointestinal tract of the enzyme fatty acid amide hydrolase¹⁰, which catalyses the hydrolysis of both compounds. Therefore, it is unlikely that the amounts of anandamide and 2-AG found in food are sufficient to produce observable psychotropic effects. We did not test other unsaturated NAEs that have been suggested to enhance endogenous anandamide activity by inhibiting its inactivation², but instead assayed oleamide — for which this kind of action is also supported by pharmacological data^{5,6}. High oral doses are again necessary for activity to be observed *in vivo* (Table 1).

Our results show that the amounts of anandamide, 2-AG and oleamide in foods, including milk and cocoa, are several orders

of magnitude below those required, if administered by mouth, to reach the blood and cause observable ‘central’ effects. The assays used here provide a gross evaluation of cannabinomimetic activity, and tests monitoring more subtle behavioural changes that might be induced by low oral doses of NAEs/oleamide are needed before the relevance of these compounds to the purported mild rewarding and craving-inducing effects of cocoa can be dismissed. The presence in milk of 2-AG (which may be released from triglycerides during digestion) and oleamide also needs to be investigated.

Vincenzo Di Marzo*, **Nunzio Sepe***,
Luciano De Petrocellis†, **Alvin Berger‡**,
Gayle Crozier‡, **Ester Fride§**,
Raphael Mechoulam§

^{*}Unità per la Chimica di Molecole di Interesse Fisiologico and [†]Istituto di Cibernetica, Consiglio Nazionale delle Ricerche, Via Toiano 6, 80072 Arco Felice, Naples, Italy
e-mail: vdm@trinc.icimib.na.cnr.it
[‡]Nestlé Research Centre, Nestec Ltd, Vers-chez-les Blanc, 1000 Lausanne 26, Switzerland
[§]Department of Natural Products, David Bloom Centre for Pharmacy, Faculty of Medicine, Hebrew University, Jerusalem 91120, Israel

1. Devane, W. A. *et al. Science* **258**, 1946–1949 (1992).
2. di Marzo, V., Beltramo, M. & Piomelli, D. *Nature* **382**, 677–678 (1996).
3. Mechoulam, R. *et al. Biochem. Pharmacol.* **50**, 83–90 (1995).
4. Cravatt, B. F. *et al. Science* **268**, 1506–1509 (1995).
5. Mechoulam, R. *et al. Nature* **389**, 25–26 (1997).
6. Bisogno, T. *et al. Eur. J. Biochem.* **254**, 634–642 (1998).
7. Cook, H. W. *Biochemistry of Lipids, Lipoproteins and Membranes* 141–169 (Elsevier, Amsterdam, 1991).
8. Martin, B. R. *et al. Pharmacol. Biochem. Behav.* **40**, 471–478 (1991).
9. Fride, E. & Mechoulam, R. *Eur. J. Pharmacol.* **231**, 313–314 (1993).
10. Katayama, K. *et al. Biochim. Biophys. Acta* **1347**, 212–218 (1997).

Beltramo and Piomelli reply — The two most abundant *N*-acylethanolamines (NAEs) found in chocolate (*N*-oleylethanolamine and *N*-linoleylethanolamine) do not activate brain cannabinoid receptors but effectively inhibit anandamide degradation^{1,2}. This observation suggested to us that these compounds might contribute to the hedonic properties of chocolate by causing non-metabolized anandamide to accumulate at its sites of action².

How can this hypothesis be tested? One possibility is to determine whether antagonists of cannabinoid receptors prevent or reduce chocolate craving. However, because cannabinoid antagonists also reduce the intake of other palatable foods and drinks^{3–5}, the results of this experiment would be difficult to interpret. Another approach would be to test whether *N*-oleylethanolamine and *N*-linoleylethanolamine act as reinforcing stimuli in behavioural experiments. It is possible, for example, to determine in animals whether these compounds possess discriminative properties, a measure of their ‘subjective’ effects. In either case, the aim is

Table 1 Effects of test substances on performance of behavioural tests in mice

	Olive oil (vehicle)	Anandamide (300 mg kg ⁻¹)	2-AG (400 mg kg ⁻¹)	Δ^9 -THC (90 mg kg ⁻¹)	Oleamide (200 mg kg ⁻¹)
Ambulation	77 ± 11	51 ± 8*	13 ± 9**	34 ± 7*	39 ± 6**
Rearing	29 ± 5	17 ± 4*	3 ± 2**	3 ± 3**	12 ± 2**
Immobility	22 ± 7	70 ± 20*	157 ± 19**	146 ± 18**	85 ± 19*
Analgesia	12 ± 0.4	16 ± 3	10 ± 1	25 ± 2**	15 ± 3
Hypothermia	-0.2 ± 0.25	-2.30 ± 0.42**	-3.2 ± 0.5**	-4.1 ± 1.0**	-2.8 ± 0.56**

Effects of orally administered olive oil, anandamide, 2-AG, Δ^9 -THC and oleamide on mouse ambulation (number of square crossings) and rearing (number of offspring reared) in an open field, immobility (time spent motionless on a ring, in seconds), analgesia (response latency to hind paw lick on a hot plate (54 °C), in seconds), and hypothermia (decrease in rectal temperature, in °C)^{6,7,9,10}. The procedures are not harmful to mice and meet with NIH ethical standards. Substances were administered to Sabra female mice by mouth by gavage. Doses lower than those shown did not produce any statistically significant effect. Results are mean ± s.e.m. from 4–9 mice. *, $P < 0.01$; **, $P < 0.005$; assessed by analysis of variance followed by Newman-Keuls post-hoc tests for group differences.

to ascertain what component of the psychopharmacological effects of chocolate, if any, is mediated by an indirect activation of the endogenous cannabinoid system.

Di Marzo and co-workers have tested the role of endogenous cannabinoids in chocolate craving by investigating the extent to which NAEs and anandamide reach the bloodstream and exert their pharmacological effects after oral administration. To test whether anandamide or oleamide (which is not an NAE, but inhibits the enzymatic hydrolysis of anandamide) elicit overt cannabis-like effects in mice, they gave anandamide or oleamide orally in quantities similar to those found in chocolate, and assessed their psychotropic effects by a standard behavioural procedure that is used in the screening of cannabinoid-receptor ligands. From their results, they conclude that the content of NAEs and other cannabinoid-related compounds in cocoa is insufficient to elicit cannabis-like effects.

These results lend themselves to two considerations. First, it seems illogical to presume that compounds present in chocolate should display a pharmacological profile similar to that of cannabis. This would be the same as assuming that cocoa and cannabis have comparable psychoactive effects, which is evidently not the case. It would have been more informative to compare anandamide and oleamide to cocoa in a standard drug discrimination test. Second, the effects of *N*-oleylethanolamine and *N*-linoleylethanolamine were not investigated: not only are these NAEs present in chocolate in greater amounts than anandamide, but they are also produced in neurons through an activity-dependent mechanism that is similar to the one implicated in anandamide formation⁶. The possibility that these compounds act synergistically to prevent anandamide degradation *in vivo* therefore remains to be investigated.

In conclusion, although the results of Di Marzo *et al.*'s study will reassure manufacturing companies that the risks of chocolate consumption do not include cannabis-like intoxication, they provide little new information on the intriguing psychopharmacology of cocoa. This substance remains, in R. J. Huxtable's apt words, "more than a food but less than a drug"⁷.

Massimiliano Beltramo, Daniele Piomelli

Department of Pharmacology, 360 Med Surge II,
University of California at Irvine,
Irvine, California 92697-4625, USA

- Desarnaud, F., Cadas, H. & Piomelli, D. *J. Biol. Chem.* **270**, 6030–6035 (1995).
- di Tomaso, E., Beltramo, M. & Piomelli, D. *Nature* **382**, 677–678 (1996).
- Arnone, M. *et al. Psychopharmacol.* **132**, 104–106 (1997).
- Colombo, G. *et al. Life Sci.* **63**, PL113–PL117 (1998).
- McGregor, I. *et al. Proc. Int. Cannabinoid Res. Soc. Meet.* **76** (1997).
- Di Marzo V. *et al. Nature* **372**, 686–691 (1994).
- Huxtable, R. J. *Nature* **382**, 411–412 (1994).

Honeybees link sights to smells

It is common for a smell or a sound to trigger a vivid recollection of an associated event in the past, even if it involves a different sensory modality and the episode occurred a long time ago^{1,2}. The human brain displays impressive cross-modal associative recall. Here we investigate whether this capacity extends to insects. We find that trained honeybees can recall a specific colour when they encounter a particular scent.

To investigate whether recall of a colour can be triggered by exposure to a scent, the two kinds of stimulus should be presented sequentially. Furthermore, the colour should not be present when the scent is encountered, and vice versa. We met these requirements by using the apparatus shown in Fig. 1a. As the bees entered the first part of the apparatus, chamber A, they received an olfactory stimulus from a vial placed at the entrance. They then had to fly to chamber B, which has two exits, one marked with the colour blue and the other with yellow. The bees had to choose the yellow exit if they had encountered the scent of mango at the entrance to chamber A, and the blue exit if they had encountered the smell of lemon. The blue and yellow labels in chamber B were interchanged every 10 minutes, and the reward moved with the appropriate colour, to make sure the bees found the reward by associating each scent with the appropriate colour, rather than with a particular exit in chamber B (left or right).

After one day's training, the bees were tested in the apparatus one at a time. When the bees encountered the scent of mango at the entrance to chamber A, they showed a strong and significant preference for the

yellow exit. When the entrance was lemon scented, they preferred the blue exit (Fig. 1b, experiment 1). The scent of mango evidently evoked recall of yellow, and lemon triggered recall of blue. Bees could also be trained to make the opposite associations: yellow with lemon, and blue with mango (Fig. 1b, experiment 2).

Bees trained on this task sometimes hesitate to choose a colour in chamber B, then return to the entrance of chamber A and hover in front of the scented vial with extended antennae, as if to sample the scent once more, before returning to chamber B to make their choice. We also found the reverse phenomenon, that bees can be trained to recall a specific scent when they see a particular colour (data not shown).

'Associative learning' describes the process by which an organism learns to associate a particular sensory stimulus with a reward or a punishment^{3–9}. The associative recall that we describe here, however, is more elaborate than this. It involves both the grouping together of reward and signalling sensory stimuli of different modalities in memory, and the recall of one member of a grouped pair when the other member is encountered.

For a foraging honeybee, cross-modal associative recall can facilitate the search for a food source; for example, detecting the scent of lavender can initiate a search for purple flowers. Indeed, one might surmise that the nectar samples received by a potential foraging recruit from a dancing bee may stimulate the recruit into recalling many of the host flower's other attributes, such as its colour, shape and texture, and perhaps even the route to the flower patch if the recruit has previously visited it.

In nature, the smell and appearance of flowers are often (but not always) detected at the same time. In our experiments we presented the olfactory and visual stimuli

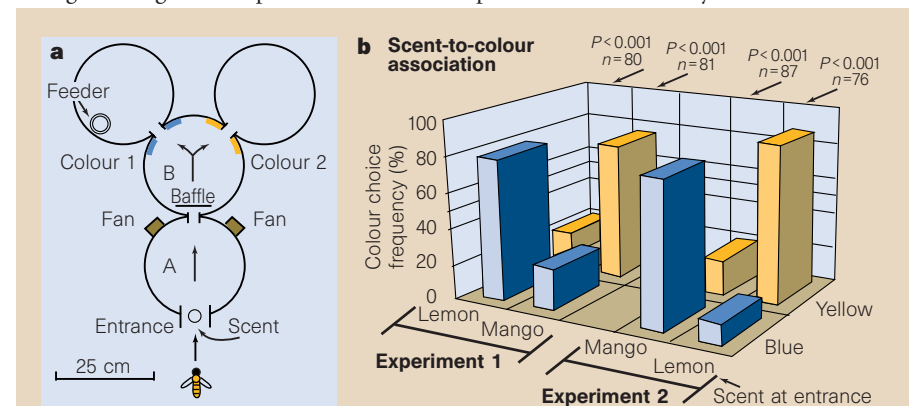


Figure 1 Cross-modal associative recall in honeybees. **a**, Experimental set-up. Bees had to choose the appropriate colour in chamber B (blue or yellow) according to the scent (lemon or mango) they experienced from a vial at the entrance to chamber A. The fans created a slight positive pressure in chamber A, ensuring that the bees encountered the scent only as they entered that chamber, and not in chamber B. The baffle ensured that the bees did not see the colours until they entered chamber B. **b**, Results of tests; *n* indicates the number of choices analysed and *P* represents the confidence level¹⁰ for choice frequencies being significantly different from random choice (red line).

sequentially, to ensure that we were eliciting associative recall. It remains to be seen whether bees that have experienced the colour and the scent of an object simultaneously can recall one of the object's attributes when the other is presented in isolation.

M. V. Srinivasan, S. W. Zhang, H. Zhu

Centre for Visual Science,
Research School of Biological Sciences,
Australian National University,
P.O. Box 475, Canberra, A.C.T. 2601, Australia
e-mail: m.srinivasan@anu.edu.au

1. Hintzman, D. L. *The Psychology of Learning and Memory* (Freeman, San Francisco, 1978).
2. Baddeley, A. *Your Memory: A User's Guide* 169–172 (Penguin, London, 1993).
3. Dickinson, A. *Contemporary Animal Learning Theory* (Cambridge Univ. Press, 1980).
4. Mackintosh, N. J. *Conditioning and Associative Learning* (Oxford Univ. Press, 1983).
5. Menzel, R. in *Neurobiology of Comparative Cognition* (eds Kesner, R. P. & Olton, D. S.) 237–292 (Erlbaum, Hillsdale, NJ, 1990).
6. Gould, J. L. & Gould, C. G. *The Honey Bee* (Freeman, New York, 1988).
7. Bitterman, M. E. *Anim. Learn. Behav.* **24**, 123–141 (1996).
8. Wehrer, R. in *Handbook of Sensory Physiology* Vol. 7/6C (ed. Autrum, H.) 287–616 (Springer, Berlin, 1981).
9. Couvillion, P. A. & Bitterman, M. E. *Anim. Learn. Behav.* **16**, 67–74 (1988).
10. van Hateren, J. H., Srinivasan, M. V. & Wait, P. B. *J. Comp. Physiol. A* **167**, 649–654 (1990).

Immobile plasticizer in flexible PVC

Plasticized poly(vinyl chloride) (PVC) is one of the most widely used polymeric materials in medical and related applications, and usually contains up to 40 per cent di-(2-ethylhexyl)phthalate (DEHP), which acts as the 'plasticizer' to impart flexibility to an otherwise rigid PVC¹. The plasticizer can migrate from PVC-based devices and storage bags into physiological fluids, however, and has been detected in storage media such as blood, plasma, serum, drug solutions and fatty foods^{2–4}, as well as in the bodies of patients undergoing haemodialysis and transfusion⁵. This is a concern because DEHP is a lipid-removing liver carcinogen⁶, and causes hepatic⁷ and reproductive toxicity⁸ in rodents, although opinion is divided on its toxicity in humans⁹.

Attempts to prevent this migration have included coating, grafting or blending PVC with other polymers, glow-discharge treatment of the surface of PVC, prolonged ultraviolet irradiation, and photocrosslinking of dithiocarbamated PVC¹⁰. We have developed a surface-modification technique to prevent this migration, and show that plasticized PVC becomes extremely resistant to migration when treated with sodium sulphide in the presence of a suitable phase-transfer catalyst in water. The treatment should benefit the medical and related applications of flexible PVC.

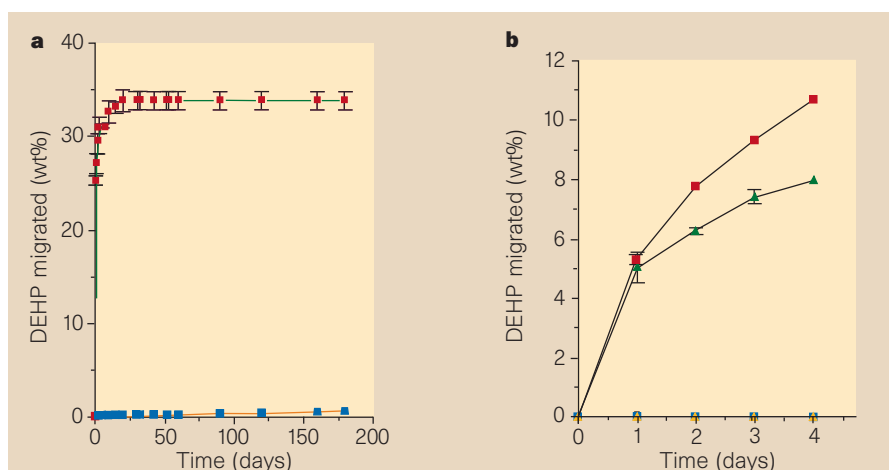


Figure 1 Surface modifications prevent migration of DEHP from PVC. **a**, Migration of DEHP from unmodified (red squares) and modified (blue squares) PVC tubes into petroleum ether at 30 °C. **b**, Migration of DEHP from unmodified (red and green symbols) and modified (blue and yellow symbols) PVC tubes into cotton-seed oil (squares) and poly(ethylene glycol)-400 (PEG400; triangles) at 70 °C.

The formation of dialkyl sulphide from an alkyl halide and sodium sulphide is well documented and proceeds readily in the presence of a suitable phase-transfer catalyst¹¹. The chlorine atoms on PVC are labile and can be substituted by various other nucleophiles¹². Our strategy was to substitute the chlorine on PVC using a dianion such as sulphide, which causes the polymer chains to crosslink. If crosslinking is confined mainly to the surface of PVC by reacting plasticized PVC in a solvent in which the polymer is insoluble, such as water, then such surface crosslinking should retard or prevent the diffusion of the plasticizer.

We therefore treated medical-grade PVC tubes (from Solmed, Denmark) with sodium sulphide (7.0 mol per litre) in the presence of tetrabutyl ammonium hydrogen sulphate (0.15 mol per litre) as the phase-transfer catalyst at 80 °C in water for 5 h, with occasional shaking. Specimens were then washed with plenty of water, sonicated for 1 min in a bath-type sonicator and dried to constant weight at 50 °C. X-ray photoelectron spectroscopy and elemental sulphur analysis showed that sulphur was present in the modified PVC specimens. The crosslinked network formed on the surface of PVC could be separated after treating it with tetrahydrofuran, causing only the uncrosslinked part to dissolve, along with other additives in flexible PVC.

The amount of DEHP that migrated into petroleum ether, a potential extractant for DEHP, over 6 months at 30 °C was determined spectrophotometrically from unmodified and surface-modified PVC tubes (Fig. 1a). Virtually no plasticizer migrated into this medium from surface-modified specimens, whereas unmodified PVC lost almost all its plasticizer in a day. Accelerated migration at 70 °C, followed by monitoring the change in weight gravimetrically, indicated that no migration

occurred from the modified specimens into media of different polarities, such as cotton-seed oil and PEG400 (Fig. 1b).

Surface modification imparted a slight yellow colour to the tubes, causing optical transmittance to decrease by about 20 per cent in the 400–500 nm region of the spectrum. At 500–700 nm, the optical transmittance of the modified tubes was similar to that of the unmodified tubes (about 80 per cent).

When the surfaces of flexible PVC sheets were similarly modified, we found that plasticizer migration could be prevented completely, as it was in PVC tubes. However, modification led to a decrease in the stress (about 8 per cent) and strain (about 28 per cent) at the breaking point for PVC sheets. There was also significantly reduced adhesion of platelets and bacteria to the surface-modified sheets.

A. Jayakrishnan, S. Lakshmi

Polymer Chemistry Division,
Biomedical Technology Wing,
Sree Chitra Tirunal Institute for Medical Sciences
and Technology, Satelmond Palace Campus,
Trivandrum 695 012, India
e-mail: bmtwing@md2.vsnl.net.in

1. Blass, C. R., Jones, C. & Courtney, J. M. *Int. J. Artif. Organs* **15**, 200–203 (1992).
2. Jaeger, R. J. & Rubin, R. J. *Science* **170**, 460–462 (1970).
3. Waugh, N. W., Trissel, L. A. & Stella, V. J. *Am. J. Hosp. Pharm.* **48**, 1520–1524 (1991).
4. Castle, L., Mayo, A. & Gilbert, J. *Food Addit. Contam.* **7**, 29–36 (1990).
5. Plonait, S. L., Nau, H., Maier, R. F., Wittfoht, W. & Obladen, M. *Transfusion* **33**, 598–605 (1993).
6. Ledwith, B. J. et al. *Mol. Carcin.* **8**, 20–27 (1993).
7. Winberg, L. D. & Badr, M. Z. *Toxicol. Lett.* **76**, 63–69 (1995).
8. Davis, B. J., Maronpot, R. R. & Heindel, J. J. *Toxicol. Appl. Pharmacol.* **128**, 216–223 (1994).
9. Cadogan, D. F. *Proc. Inst. Materials Int. Conf. 'PVC 96'* (British Plastics Federation, London, 1996).
10. Lakshmi, S. & Jayakrishnan, A. *Polymer* **39**, 151–157 (1998).
11. Starks, C. M. & Liotta, C. *Phase Transfer Catalysis: Principles and Techniques* (Academic, New York, 1979).
12. Okawara, M. & Ochiai, Y. in *Modification of Polymers* (eds Carraher, C. E. & Tsuda, M.) 41–57 (American Chemical Society, Symp. ser. 121, Washington DC, 1980).

The secret life of the Neanderthal

Ancestral Images: The Iconography of Human Origins

by Stephanie Moser

Sutton: 1998. 200 pp. £25

Nicolaas A. Rupke

Boring pages of text can be brought to life by illustrations. This is an old and widespread view. Pictures — in this perception — add to the readability of verbal communication and provide additional clarity. Scientific illustrations are no exception to this rule, elucidating and animating what in many instances is dense and turgid prose. Drawings, woodcuts, engravings, lithographs and photographs have successively found their way into mediaeval illuminated manuscripts, Renaissance anatomy texts, early-modern natural philosophy books and late-modern scientific literature.

In recent years, scientific illustrations have begun to be looked at in a different way, however. Historians of science, art historians, palaeontologists, archaeologists and others have started to pay more serious attention to pictures and to regard them as not merely a means to elucidate and enliven the written word, but as a significant form of communication in their own right.

Illustrations have their peculiar iconographic idiom and communicate information that may be different from that of the accompanying text and, what is more, may enter territory that is off limits to verbal communication. Visual images, moreover, can have a powerful impact on the mind and insidiously impart cultural and other prejudices, not fit to be expressed in words, or reveal details for which no scientific evidence is available. The recognition of these and other functions of non-verbal communication in the sciences has led to the lively new discipline of 'visual communication', encompassing, in addition to illustrations, three-dimensional models and such museum displays as habitat dioramas.

One archaeologist who has taken a lead in the study of science's visual language is Stephanie Moser, who teaches at the University of Southampton. Having become sensitized to the subtle power of visual communication by an interest in the gender assumptions of archaeological reconstructions, she extended the range of her expertise to include the topic of visual representation of human prehistory in general. The fruit of this is this splendid book, which focuses on Victorian and twentieth-century scientific illustrations of the origins and early history of humans.

Among the examples Moser discusses are illustrations by Emile Bayard for Louis Figuier's *L'homme primitif* (1870), by Cecil



Naked aggression: artistic tradition has governed the way primitive peoples are depicted over a long period, as in this sixteenth-century watercolour.

Aldin for H. N. Hutchinson's *Prehistoric Man and Beast* (1896) and paintings by Zdenek Burian for his and Josef Augusta's *Prehistoric Man* (1960). We may think that our perception of human prehistory is based on excavated fossils, artefacts and other prehistoric remnants — that the classic works by T. H. Huxley, Carl Vogt, Charles Darwin or Charles Lyell on human antiquity and evolution shaped our ancestral images. Moser, however, maintains that the way we visualize 'primitive man' is significantly conditioned by artistic representations that draw on a multitude of different and pre-scientific iconographic traditions.

One important question in studying visual representation in the sciences has been that of the relative importance of 'form' and 'function'. In other words, are features of apparent scientific content not, in fact, a product of artistic style and iconographic tradition? Moser emphatically comes down on the side of the formalists. Her fascinating and challenging thesis is that late-modern, scientific visualizations of human origins retained emblematic or iconographic elements from pre-scientific traditions of representation, going back to early-modern and Renaissance, mediaeval, early Christian and even Graeco-Roman imagery — "As early as the classical period the roots of a visual tradition for depicting prehistory were established."

There was an artistic tradition of showing primitiveness that existed millennia

before Victorian palaeoanthropology. Various elements entered the visual language about primaeval humankind that served to construct a relationship of progress and of superiority of 'us' over 'them'. Among these elements were nakedness, the wearing of animal skins, hirsuteness and the presence of gnarled clubs. How many of us know that there is no factual evidence for the club "that any design-conscious Neanderthal simply must carry as an accessory"?

It might seem churlish to point out omissions in a book that offers such a rich and colourful palette of topics. Yet inasmuch as chapter one professes to present the main developments in scientific visualization, it might for the sake of completeness be appropriate to mention an oversight, namely that of the Humboldtian revolution in scientific visualization. Alexander von Humboldt's altitude–latitude representations of the distribution of vegetation, his reinvention of the isoline and his insistence on the significance of landscape perception led to a proliferation of scientific visual representations, in atlases, textbooks and so on. Ethnology and archaeology, too, were caught up in this maelstrom of Humboldtian visualization.

This is an enjoyable piece of scholarship in which the author convincingly argues that our ideas about the human past have in part been shaped by the visual arts. Moser's book is complemented by a spirited foreword by Clive Gamble, who stresses the independence from scientific discoveries of pictorial reconstruction in palaeoanthropology and points to their gendered, male-centred nature. Moser's lavishly illustrated book is accessible to a wide readership and would make a fine Christmas present. □

Nicolaas A. Rupke is at the Institute for the History of Science, Göttingen University, 37073 Göttingen, Germany.

An army of zombies

Phantoms in the Brain: Probing the Mysteries of the Human Mind

by V. S. Ramachandran and Sandra Blakeslee

William Morrow: 1998. 326 pp. \$27, £17.99

Raymond J. Dolan

A pervasive optimism is evident in the cognitive neurosciences, which is no doubt a very good thing for publishers. Like Hollywood movie studios, publishers have a distinct preference for the happy ending when it comes to science. For "hero gets the girl" read "scientist solves the mystery". Not unexpectedly, this optimism receives a resounding endorsement in *Phantoms in the Brain* by

V. S. Ramachandran, a neuroscientist, and Sandra Blakeslee, a journalist.

In all fairness, it has to be said that an upbeat tone is, almost without exception, a unifying feature in the mass of recent books concerning the human brain. The thinking behind this optimism goes as follows: understanding the human brain more or less adds up to understanding the human mind. The exponential growth in knowledge about the brain means that at the dawn of a new millennium we are close to a neural account of mental life. There is nothing inherently wrong with this except that it hides the fact that there is an enormous gulf between neurons and thought processes. We have no real idea of the computations performed by the former that enable the latter.

Phantoms in the Brain is structured around the life histories of patients with unusual, often bizarre, neurological disorders. This case-history approach has, of course, been popularized by Oliver Sacks, who contributes a foreword to the present volume. The narrative structure of biography has undoubted pedagogic value, but it is worth remembering that the worst examples of the genre come uncomfortably close to modern-day equivalents of a travelling freak-show. To their credit, this shameful pitfall is skilfully avoided by Ramachandran and Blakeslee, who keep the ideas they are seeking to communicate centre stage.

The book's overall message is that current knowledge of brain function has reached a point where even the strangest psychological and neurological disorders are amenable to neuroscientific explanation. This argument is most convincing for the phenomenon of 'phantom limbs', in which a person, following an amputation, continues to sense the limb's presence. Movement, itching or pain may still be experienced. Ramachandran claims "to have genuinely solved the mystery" of what is happening here.

The putative solution, which represents the accumulated work of many distinguished scientists, can be summarized as follows. The cortex of the human brain contains 'maps' that represent bodily sensations and movements. In the absence of sensory inputs to these maps, as is the case after limb amputation, for example, the maps undergo extensive reorganization. This rearrangement involves progressive encroachment from surrounding maps that represent other, often adjacent, body surfaces. Consequently, following an amputation, sensory experiences in an absent limb can be elicited by stimulation of body surfaces whose cortical maps have invaded the territory of the amputated region. The face and arms have adjoining cortical maps and so, not surprisingly, face stimulation can elicit sensation in a phantom arm.

The critical importance of reorganization is also attested by the fact that the intensity of phantom limb experiences is directly propor-



Mind field: can neuroscience yet adequately explain the strangest psychological disorders?

tional to the degree of cortical remodelling. The account provided in *Phantoms in the Brain* of phantom limb phenomena strikes an exemplary balance between narrative, amusing anecdote and clear scientific exposition.

Although the discussion of phantom limbs forms a substantial part of the book, there is also extensive treatment of a more elusive phantom, a zombie within, capable of making perceptual discriminations of which a person is not consciously aware. It is now well known that patients with profound perceptual impairments can nevertheless perform directed and appropriate actions towards 'unseen' environmental targets. The classic examples include blindsight, neglect and the agnosias. So, for example, in blindsight a patient who is blind in a portion of their usual field can nevertheless indicate the detection of a moving target there. The severe dissociations in these patients provide the ammunition for Ramachandran to argue that the experience of self may be just an illusion that provides a convenient mechanism for organizing the army of zombies inside one's head.

No serious account of self, or consciousness, can ignore emotion and feelings. Their relevance to a sense of self is illustrated by what is surely one of the oddest mental afflictions, Capgras' syndrome. Here a person acquires the belief that someone they know, usually a close relative or spouse, has been replaced by an identical double. In all other respects, a person with this condition has the appearance of being completely rational. Notably, they have no difficulties with visual recognition or high-level visual discrimination tasks. What is also well documented is that this type of false belief, or delusion, often develops following a brain injury.

A proposed explanation is that the neural machinery for perceptual discrimination is intact but that a parallel system for emotional evaluation of visual objects, such as people, is

dysfunctional. Consequently, the emotional resonance linked to recognition of a particular individual is blunted or absent. Ramachandran, and others, argue that a strategy for dealing with the ensuing experiential anomaly is to assert that the real person has been replaced by an identical double. Of course, this highly plausible account implies that when push comes to shove emotional evaluations hold greater sway in the theatre of the mind than rational appraisals.

This book will have wide appeal and will further whet the appetite of a public with an increasing hunger for information about the brain. But mass-market appeal is notoriously associated with compromise, and many would question whether the buzz words of the era — synthesis, coherence, consilience — convey an excessively comforting view of the nature of science. The optimistic tone that is good for book sales fails to reflect the toil, tedium and frustration that makes up much of what Thomas Kuhn referred to as normal science. This begs the question of what, other than fortune and fame, should motivate anyone to write a science book. Perhaps it should be considered a modern day virtue that is its own sufficient reward. □

Raymond J. Dolan is in the Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London WC1N 3BG, UK.

Knowing your roots

Dental Anthropology: Fundamentals, Limits and Prospects

edited by Kurt W. Alt, Friedrich W. Rösing and Maria Teschler-Nicola
Springer: 1998. 564 pp. \$89.95

Simon Hillson

Dental anthropology is the study of people from the evidence provided by their teeth. As with other areas of anthropology, 'people' is used in a broad sense to include our extinct relatives taken in the context of primates as a whole. It has been one of the less well known sub-disciplines of anthropology, but the past two years have seen no fewer than five books published on the subject, on top of several earlier texts. Dental papers are regularly published in the main anthropological journals, and there is a thriving Dental Anthropology Association with more than 200 active members and its own journal.

Why teeth, and why the growth in interest? Teeth and bones are the main research material for anthropologists who study the fossil remains of hominids and other primates, human remains from archaeological sites, and forensic cases. There are particular advantages in studying teeth, however. They survive better in the ground because they are tougher and more heavily mineralized than bone, and they are among the most common fossils found. Teeth provide a protected



Ancestral diet: *Australopithecus anamensis* ate fruit and nuts, according to the evidence of teeth found in Kenya by Meave Leakey and her team.

environment for the survival of biochemical information and microscopic detail. Once formed in childhood, their component tissues do not turn over like bone, so the sequence of formation is preserved inside them as layered structures. Scattered teeth from one individual can be matched up by this layering pattern, and it is possible to reconstruct the timing of formation. Dental development is known from X-ray studies of living children and, as the sequence is less variable than bone development, it is the best method for estimating age at death in children's remains.

Once in the mouth, teeth are exposed to all food that enters the body, and they are used for many purposes besides eating. They were an important part of the toolkit of recent hunter-gatherers, whose lifestyle left its mark in the pattern of dental wear and the epidemiology of dental caries and periodontal disease. Strong contrasts are shown between hunter-gatherers and early agriculturalists, and with later townspeople. Teeth therefore provide some of the best evidence for reconstructing the life of a person from their fragmentary remains. It is possible to estimate age at death in adults, from the state of wear in the rapidly wearing teeth of past populations, or from age-related histological changes in the roots of modern forensic cases. Tooth size varies between men and women and, because teeth are formed adult-size, this can be used to distinguish girls from boys where the skeleton is insufficiently developed to yield many clues.

Tooth size also figures in the broader scale of human evolution — dental reduction is one of the most prominent evolutionary changes over the past 30,000 years. There is

also variation in the form of cusps, roots and other features which appear with characteristic frequencies in broad geographical groupings of living populations, and provide important evidence for the way in which modern humans evolved.

Dental Anthropology: Fundamentals, Limits and Prospects has broad aims. Its editors' stated intention is to provide an introduction as well as a reference for specialists, and they hope that it will contribute to further research. The extent to which all these aims are met is determined by its format as an edited collection of papers by many authors. Several papers take independent and most interesting lines, which are thought provoking for specialists and do suggest directions for research. They will also be a good source for advanced student essays and seminars, but they are not really introductory. One very valuable feature is the copious reference material, including literature that is less frequently cited in mainstream anthropology journals. But some chapters have puzzling omissions which a student or general reader would need to make good from other sources. In among the more specialist contributions, there are also chapters that are clearly designed as an introduction, such as the excellent concise history of dental anthropology.

Overall, most areas of dental anthropology outlined above are covered, although not all in the same detail. Variation in tooth form, for example, is relatively underplayed, although many would see it as a core area of dental anthropology. Many features not often discussed elsewhere are described, but only a brief mention is made of the main features of recent population studies.

I enjoyed reading this book very much. It includes material that is not available elsewhere, and I will use it as a valuable reference.

Specialist dental anthropologists will wish to own a copy, and forensic scientists and clinical dentists will find it useful. I will also recommend it to my students. □

Simon Hillson is at the Institute of Archaeology, University College London, 31–34 Gordon Square, London WC1H 0PY, UK.

Far beneath in the abysmal sea

The Search for the Giant Squid

by Richard Ellis

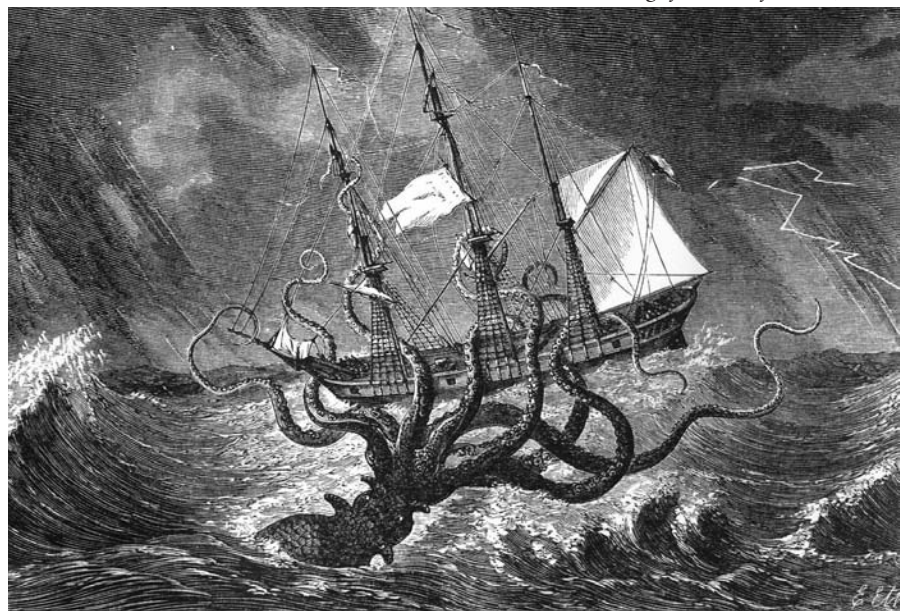
Lyons Press: 1988. 316 pp. \$35

Martin Wells

"It's as big as a city bus — and it's real" starts the press release, in bold print. Oh dear. Another gee-whiz product at the lower end of journalism. I almost refused to look at the book, let alone agree to review it.

Thank goodness I didn't turn it down. It isn't that sort of book at all — Richard Ellis should have a few unfriendly words with his publisher's publicity department. Because what he has written — and painted, his own reconstructions ring true — is serious and very well researched. Ellis has collected together the scientific evidence as well as the flights of fancy that have accumulated around this largest of all invertebrates.

His problem, of course, is what people would like to believe. "Men really need sea-monsters in their personal oceans. For the ocean, deep and black in the depths, is like the low dark levels of our minds in which the dream symbols incubate and sometimes rise up to sight like the Old Man of the Sea. An ocean without its unnamed monsters would be like a dreamless sleep" (Ellis quotes John Steinbeck's *The Log of the Sea of Cortez*).



"The Kraken, as seen by the eye of the imagination": giant squid have featured in many sailors' tales, as seen here in an illustration from John Gibson's *Monsters of the Sea* (1887).

Many of us would be saddened if we really have already hit upon giant squid as big, or almost as big, as they get. Not that an animal proven to reach a length of 15 or 20 feet, with tentacles extending to 50, is by any means negligible; it is just that, deep down, we would like to believe that there are creatures the size of submarines waiting to grab at us in the unexplored wilderness of the deep sea.

So Ellis sets out to keep us all happy, by quoting sailors' tales, many — but let's hope not all of them — gross exaggerations, as well as taking a hard-headed look at the evidence of the specimens washed up on beaches, trawled from the depths or recovered from the stomachs of sperm whales. Sperm whales are a lot better at catching these creatures than we are ("we only catch the slow, the sick, and the stupid" — again a quote, this time from Clyde Roper, at the Smithsonian Institution in Washington DC). Fortunately, sperm whales have difficulty digesting the beaks of squid, so we know something of the sizes, distribution and probable numbers of *Architeuthis* from their gut contents — a source of information, incidentally, that has dried up with the moratorium on whaling.

It seems that giant squid (*Architeuthis* is not the only genus of big squid, although it seems to include all the real monsters) are very widely distributed, and probably quite common. Sperm whales seem to eat an awful lot of them and, at least when adult, the big squid prefer cold, deep waters. So the ones we find at the surface are perhaps always moribund.

The sad fact emerges that nobody has ever seen a giant squid in good condition. And most of the available evidence suggests that the animals are probably relatively slow moving, drifters rather than drivers. They are neutrally buoyant, because much of their muscle tissue is replaced by cells retaining ammonium chloride — which is lighter than sea water and, incidentally, makes them thoroughly unpalatable to us. The two tentacles responsible for most of the total length of the giants are probably far too long and thin to be shot out at prey in the manner of the cuttlefish and smaller squid that we observe in aquaria. There is a surprising lack of the giant nerve fibres that would ensure that the mantle contracts simultaneously all over, and the cartilage locking the mantle to the funnel is poorly developed, so that a really powerful jet is perhaps beyond them.

Ellis's book tells us all this, as well as listing all the authenticated sightings and strandings of *Architeuthis*, and reviewing the history of the many sea-serpents that were probably, in fact, squid. There are chapters on the giant squid in literature and in films and about the several reconstructions of giant squid made for museums throughout the world — Hitler, he reminds us, destroyed the model in the Natural History Museum in London.

There are plenty of illustrations from imaginative accounts of contacts in the past

and a liberal sprinkling of photographs of men in white coats standing around (and usually somewhat behind — even scientists in white coats really want squids to be big) the laid-out remains of more recent encounters. A gold-mine of fact and fantasy, for we scientists who work on cephalopods and for all of us who love monsters. □

Martin Wells is in the Department of Zoology, University of Cambridge, Cambridge CB2 1TN, UK.

The cosmic rain

The Particle Century

edited by Gordon Fraser

Institute of Physics Publishing: 1998. 227 pp.

\$59.50, £30

David Miller

Among the authors of these 20 essays there are sufficient 'good communicators' that the serious reader can learn a great deal about the science, and there are sufficient 'great names' to give a thrill of authenticity from the first-hand accounts of discoveries. And some of the 'great names' are among the best communicators. The story really starts with two theoretical insights, Paul Dirac's prediction of anti-electrons and the assertion by Hideki Yukawa that the short-range force that binds protons and neutrons into nuclei must be generated by the exchange of a massive particle.

Both were confirmed by observations of cosmic ray interactions: the positron (the anti-electron) very soon after its prediction; the pion (Yukawa's force carrier) only after a long delay from 1935 to 1947, despite the early discovery of a 'mesotron' with many — but not all — of the right properties. This later turned out to have been the muon — the first heavy copy of the electron, as explained in the essay on the "Cosmic Rain" by Owen Lock, who began his PhD work at Bristol just after Cecil Powell had identified the pion as a separate object from the muon.

Cosmic rays also generated the first handfuls of strange particles, but they were only properly understood after accelerators generated hundreds of them, and bubble chambers pictured their behaviour. Nick Samios writes about the quest to observe the whole set of strange particles which culminated in his team's discovery in 1964 of the omega-minus, whose properties had already been predicted from SU(3) theory. One of those who made the prediction was Yuval Ne'eman, who tells how falteringly the idea of quarks grew from the mathematical structure of SU(3) into a physical hypothesis. Physicists only began to believe that the three quarks of SU(3) (up, down and strange) were real after seeing deep inelastic scattering of electrons and neutrinos from pointlike objects inside protons — described here by Jerry Friedman and Henry Kendall, two leaders of the team who did the first such experiment. □

That was the beginning of the revolution of the early 1970s, when the whole structure of the present Standard Model began to fall into place, driven by a rapid dialogue between experimental surprises and theoretical hypotheses. Roy Schwitters describes how the discovery of a spin 1 resonance, the ψ/J , was matched to a theory of weak interactions which needed a fourth (charm) quark which could bind to its own antiquark to make just such a spin 1 state. Once charm was understood it became possible to unify the weak and electromagnetic theories into a single gauge-field theory — described in a superbly lucid essay by Martinus Veltman, one of its founders. This 'electroweak' theory needed the W and the Z bosons as its force carriers. They are about 100 times heavier than a proton — which makes them very hard to produce. Carlo Rubbia explains in his quirky English how he and colleagues built a machine to collide protons and antiprotons at sufficiently high energy to prove that the W and Z existed. At the same time a gauge-field theory of the strong interaction between quarks developed, Quantum Chromodynamics (QCD), discussed by Sau Lan Wu, one of the team that proved the existence of the gluon, the gauge-boson which carries the strong force.

There are healthy overlaps between authors. If a reader does not understand a concept on first encounter — say the need for a top quark in Guido Altarelli's essay on how the Standard Model works — it might become clearer after reading Mel Shochet. Some essays are straight story-telling, some round out the picture by explaining the theory, or how the accelerators and detectors work. Even at the end of the century there are gaps in the Standard Model: is there a Higgs boson, how do we fit gravity in, why are some symmetries almost but not quite exact? As John Ellis points out, these gaps link up with profound questions in cosmology.

The first essay is different from all the rest, and extremely valuable. Catherine Westfall and John Krige give a brief social and political history of post-war physics. They show, among other things, how in the United States the rivalries of the Cold War drove science along, while in Europe science was a neutral ground on which enmities could be forgotten, leading to the still growing success of the European Laboratory for Particle Physics (CERN) in Switzerland. And there are warnings about how governments now want to 'acquire' science rather than to work in partnership with the scientist. This book could do a little to help restore that partnership. The dedicated non-scientist will get a lot from it, although its main audience is likely to be among school and undergraduate students of physics. □

David Miller is in the Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK.

Genetic instabilities in human cancers

Christoph Lengauer, Kenneth W. Kinzler & Bert Vogelstein

Whether and how human tumours are genetically unstable has been debated for decades. There is now evidence that most cancers may indeed be genetically unstable, but that the instability exists at two distinct levels. In a small subset of tumours, the instability is observed at the nucleotide level and results in base substitutions or deletions or insertions of a few nucleotides. In most other cancers, the instability is observed at the chromosome level, resulting in losses and gains of whole chromosomes or large portions thereof. Recognition and comparison of these instabilities are leading to new insights into tumour pathogenesis.

It is now widely accepted that cancer results from the accumulation of mutations in the genes that directly control cell birth or cell death. But the mechanisms through which these mutations are generated are the subject of continuing debate. It has been argued that an underlying genetic instability is absolutely required for the generation of the multiple mutations that underlie cancer^{1,2}. However, it has also been suggested that normal rates of mutation, coupled with waves of clonal expansion, are sufficient for the process to occur in humans³. Here we review observations relating to the stability of the genome of human cancer cells. Although these observations do not prove that genetic instability is necessary for a tumour to develop, they strongly support the existence of two levels of instability, one or the other of which is affected in the vast majority of cancers.

Types of genetic alterations in tumours

Numerous genetic alterations that affect growth-controlling genes have been identified in neoplastic cells over the past 15 years, providing persuasive evidence for the genetic basis of human cancer. The alterations can be divided into four major categories (Fig. 1).

Subtle sequence changes. These changes involve base substitutions or deletions or insertions of a few nucleotides (Fig. 1a, b), and, unlike the alterations described below, they cannot be detected through cytogenetic analysis. For example, missense mutations in the *K-ras* gene occur in over 80% of pancreatic cancers⁴.

Alterations in chromosome number. Alterations in chromosome number involve losses or gains of whole chromosomes (aneuploidy; Fig. 1c). Such changes are found in nearly all major human tumour types⁵. Examples include the loss of chromosome 10 in glioblastomas, often reflecting the inactivation of the tumour-suppressor gene *PTEN*⁶, and the gain of chromosome 7 in papillary renal carcinomas, reflecting a duplication of a mutant *MET* oncogene⁷.

Chromosome translocations. These alterations can be detected cytogenetically as fusions of different chromosomes or of normally non-contiguous segments of a single chromosome (Fig. 1d). At the molecular level, such translocations can give rise to fusions between two different genes, endowing the fused transcript with tumorigenic properties. An example is provided by the Philadelphia chromosome in chronic myelogenous leukaemias; the carboxy terminus of the *c-abl* gene on chromosome 9 is joined to the amino terminus of the *BCR* gene on chromosome 22 (ref. 8).

Gene amplifications. Gene amplifications are seen at the cytogenetic level as homogeneously stained regions or double minutes (Fig. 1e). At the molecular level, multiple copies of an 'amplicon' containing a growth-promoting gene(s) can be seen. The amplicons contain 0.5–10 megabases of DNA, and are different from the duplications of much larger chromosome regions that result from aneuploidy and translocations⁹. An example is the amplification of

N-myc that occurs in ~30% of advanced neuroblastomas¹⁰.

Differences between 'state' and 'rate'

All four of the alterations described above occur commonly in specific tumour types and are rarely or never observed in normal cells. But the existence of genetic alterations in a tumour, even when frequent, does not mean that the tumour is genetically unstable. Instability is, by definition, a matter of rate, and the existence of a mutation (state) provides no information about the rate of its occurrence. Several factors in addition to a true instability could explain a higher prevalence of mutations in tumours compared with normal cells. The selective conditions in the tumour environment, including aberrant humoral, cell–substratum and cell–cell interactions, are different than those in the environment of normal cells. As a result, mutations in growth-controlling genes could occur at the same rate in tumour cells as in normal cells, but the selective advantage posed by such a mutation in the tumour cells' environment would give rise to clonal expansion, allowing the tumour cell with the mutation to overtake its sister cells. The same mutation occurring in a normal cell, in the absence of clonal expansion, would be undetectable, as it would be masked by millions of sister cells without the mutation. It is also possible that normal cells undergoing mutations are destroyed by apoptosis, a pathway leading to programmed cell death that may protect the organism from the tumorigenic consequences of such mutations. Indeed, mutations in several oncogenes lead to growth stimulation under some circumstances but to apoptosis under others (see, for example, ref. 11).

Microorganisms can be used to illustrate the differences between mutation rate and state. Consider two yeast strains, A and B, each of which initially contains a wild-type gene, *X*. A and B are grown separately in different media. After growth for an equivalent number of generations, it is found that a mutation of gene *X* is present only in A cells. It cannot be concluded that the rate of mutation of gene *X* is higher in A cells than in B cells. The selective conditions operative in the media in which A cells were grown might simply have favoured growth of cells with mutations in gene *X*, even though such mutations arose at equal rates in both cell types. Another alternative is that strain A had a pre-existing mutation in another gene that allowed it to survive in the presence of mutant gene *X*, just as the multiple mutations found in tumours can function cooperatively.

A further complication results from the dynamic nature of tumours. Cancers generally take decades to develop, and it is important to determine whether a genetic instability, even if present at some stage of tumorigenesis, persists. As it is nearly impossible to exclude the occurrence of a transient state of instability during an undefined period of a tumour's history, we will restrict our discussion of instabilities to those that endure throughout the lifetime of

the tumour cell and which can therefore be analysed experimentally.

Subtle sequence instabilities

The first type of instability to be quantitatively assessed in neoplastic cells involved subtle sequence changes that alter one or a few base pairs. Instabilities of this type are uncommon in human cancers, but, when present, they cause dramatic phenotypes.

The molecular mechanisms underlying the fidelity of DNA replication have been summarized^{12,13}. Two separate processes control the error rate—DNA polymerization (and associated proof-reading by the polymerases) and repair. The repair machinery operates on sequence errors generated by polymerases or by mutagens. No consistent pattern of defects in polymerases has been seen in tumours. However, defects in two of the major repair systems—nucleotide-excision repair and mismatch repair—have been well documented.

Nucleotide-excision repair. Nucleotide-excision repair (NER) is responsible for repairing damage caused by many exogenous mutagens. The role of NER in cancer first came to light through

the study of xeroderma pigmentosum and related disorders¹⁴. Patients with these autosomal recessive, inherited diseases develop numerous skin tumours in sun-exposed areas. Fibroblasts from xeroderma pigmentosum patients are sensitive to ultraviolet light, and the patterns of such sensitivity, as well as the sensitivities found in fusions of cells from different patients, indicated that several different genes were likely to be involved¹⁵. This prediction was confirmed through biochemical and genetic experiments which eventually led to the identification of the NER genes, mutations in which result in these disorders¹⁶.

The biochemistry of NER has been reviewed elsewhere¹⁷, and here we focus on the relationship of NER defects to neoplasia. Skin tumours represent the major tumour type to which patients with NER defects are susceptible; the incidence of internal cancers in patients with NER defects is not raised to the same degree^{18–20}. This is surprising, because NER can correct covalent modifications of DNA resulting from various chemical mutagens, and fibroblasts from xeroderma pigmentosum patients are more sensitive than normal cells to such compounds¹⁶. The simplest explanation for these results is that ultraviolet light is the major mutagen that results in NER-correctable DNA damage to which humans are exposed^{18–20}. The interpretation offers support for the idea that environmental agents (other than ultraviolet light) are not rate-limiting for the development of most common cancers²¹.

It is interesting that xeroderma pigmentosum heterozygotes are at no higher risk for neoplasia than the general population¹⁶. The reason for this is discussed below.

Mismatch repair. Mismatch repair (MMR) was discovered in prokaryotes long ago but has been shown to be involved in cancer only within the past five years. The first clue to the role of MMR in cancer came with the discovery of a group of sporadic (non-familial) colorectal tumours that exhibited widespread alterations of poly(A) tracts in their genomes^{22,23}. Subsequent studies showed that other such 'microsatellites', including poly(CA) repeats, were similarly affected, giving rise to the term 'microsatellite instability' (MIN)^{23,24}. Although it is unusual in sporadic colorectal cancers (CRCs), MIN occurs in most cancers in patients with hereditary non-polyposis colorectal cancer (HNPCC)²⁵. The loci responsible for HNPCC were linkage-mapped to chromosomes 2p16 and 3p21 (refs 26, 27). These results indicated that HNPCC patients might inherit a replication or repair gene that was somatically mutated in the sporadic CRC cases associated with MIN, and that two such genes may reside on human chromosomes 2 and 3 (refs 23, 25–27).

None of the initial reports on MIN proposed a specific repair system that might be responsible for the instability. Strand *et al.*²⁸ were the first to suggest that this phenotype might result from defective mismatch repair: they noted that microsatellite instability had been observed in bacteria with defects in the mismatch-repair genes *mutS* or *mutL*, and showed that *Saccharomyces cerevisiae* with defects in the yeast homologues of either *mutS* or *mutL* exhibited a similar MIN phenotype. This hypothesis²⁸ was confirmed by the identification of the *mutS* homologue *hMSH2* on human chromosome 2 (ref. 29) and the demonstration of inactivating mutations in this gene in chromosome-2-linked HNPCC kindreds³⁰. A *mutL* homologue (*hMLH1*) on chromosome 3 was found to be mutant in the germ line of chromosome-3-linked HNPCC kindreds^{31,32}. This genetic evidence was strongly supported by studies showing that tumours exhibiting MIN lacked detectable MMR activity in biochemical assays^{33,34}. Six human *mutS* or *mutL* homologues are now known which, when inactivated by mutation, lead to a MIN phenotype in cancer patients³⁵.

The biochemistry of MMR has been reviewed^{36,37}, so we focus here on what has been learnt about neoplasia from studies of defects in MMR. Perhaps the most important of these lessons concerns the dominant inheritance of HNPCC. In general, patients with HNPCC have one normal allele of the relevant MMR gene; this wild-type allele is sufficient to maintain normal levels of MMR³³. It is only

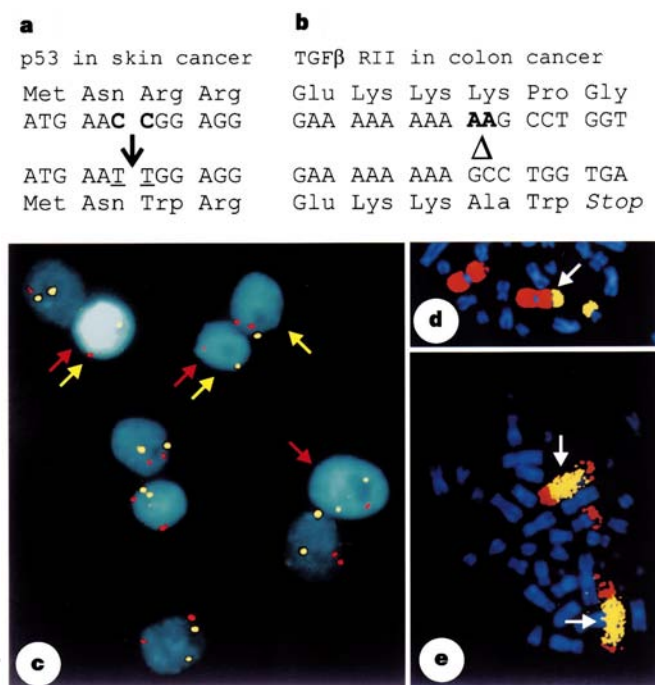


Figure 1 Examples of genetic alterations in cancer. **a**, **b**, Subtle sequence alterations: **a**, mutation at a dipyrimidine site (bold letters) of the *p53* gene (codons 247–248) found in a xeroderma pigmentosum patient with a defect in nucleotide-excision repair (NER)³⁴; **b**, a two-base deletion located within a sequence of ten repeating adenines of the transforming growth factor-β receptor II (TGFβ RII) gene (codons 125–128) in a colorectal cancer cell line with mismatch-repair (MMR) deficiency³⁵. **c**, Gross chromosomal change. Loss of chromosomes 3 (red arrows) and 12 (yellow arrows) in colorectal cancer (CRC) cells. A clone of the CRC cell line SW837 was expanded through 25 generations before fluorescence *in situ* hybridization (FISH). Interphase nuclei were hybridized with labelled centromeric DNA probes specific for chromosome 3 (red spots) and chromosome 12 (yellow spots). The number of signals detected in SW837 cells was diverse, indicating CIN; normal cells, as well as cancer cells exhibiting MIN, had two red and two yellow signals in nearly every nucleus⁴⁶. **d**, Chromosome translocation. A metaphase plate of the neuroblastoma cell line GIMEN was hybridized by FISH with labelled whole-chromosome-painting probes specific for chromosome 1 (red) and chromosome 17 (yellow), revealing a t(1;17) translocation (arrow). **e**, Gene amplification. FISH with a *N-myc* probe (yellow) and a whole-chromosome-painting probe specific for chromosome 1 (red) revealed an area of *N-myc* amplification (arrow) within the derivative chromosomes 1 of the neuroblastoma cell line Kelly.

when the wild-type allele is inactivated during tumorigenesis—through a gross chromosomal event or a subtle intragenic mutation—that MMR is abolished and mutations accumulate (Fig. 2). The progeny of this cell then accumulate mutations in oncogenes and tumour-suppressor genes, thus resulting in clonal expansion (that is, tumorigenesis). These results and analogous MMR gene mutations in sporadic tumours unambiguously show that somatic mutations in repair genes can be selected for during tumorigenesis, even when such mutations do not directly enhance cell growth.

Although patients heterozygous for a MMR-gene mutation are tumour-prone^{30,35}, patients heterozygous for a NER-gene mutation are not¹⁶. One possible explanation for this difference involves the extra step required to lead to the accumulation of mutations in the NER-gene heterozygotes. In a heterozygote with one defective MMR allele, all that is required to begin to accumulate mutations at a high rate is the inactivation of the normal allele inherited from the unaffected parent. This inactivation immediately leads to uncorrected errors during normal DNA replication. In a heterozygote with one defective NER allele, however, inactivation of the normal allele does not always lead to a high mutation rate. Exposure to an environmental agent (that is, ultraviolet light) is required to engender such mutations (Fig. 2).

How frequently do MMR defects occur in human neoplasms? Roughly 13% of colorectal cancers and a similar percentage of endometrial and gastric cancers contain such defects³⁸. Cancers of other types are rarely (<2%) MMR-deficient. Although many cancers may contain a few microsatellite changes, the extent of these changes is much less pronounced than in MMR-deficient cancers. These low levels of microsatellite alterations are likely to reflect near-normal rates of mutation during the multiple rounds of

cellular division that accompany the waves of clonal expansion through which advanced cancers evolve³⁹.

Do cancers that do not exhibit MMR or NER deficiency have other replication or repair abnormalities that render them prone to subtle mutations? The answer is, probably not. Assays that detect such subtle mutations (base substitutions or small deletions or insertions) have been used to study several tumour-cell lines. The mutation rates in such lines have, in general, been found to be roughly equal to those in normal cell lines, with the exception of tumour lines with reported defects in MMR or NER^{1,18}. However, it is difficult to compare mutation rates among tumour and normal cell lines, which grow and age at different rates and are often grown under different conditions. A more rigorous comparison can be made between MMR-proficient and MMR-deficient colorectal cancer cell lines grown under identical conditions. There is a dramatic difference in the mutation rates between these two types of cancer, with the MMR-proficient tumours exhibiting mutation rates similar to those observed in normal cells^{40,41} (Fig. 3).

Chromosome number instabilities

The subtle sequence instabilities represented by NER-associated instability (NIN) and MIN are rare, but another form of instability, involving gains and losses of whole chromosomes, is likely to occur in most human malignancies⁵. Karyotypic studies have shown that the majority of cancers have lost or gained chromosomes, and molecular studies indicate that karyotypic data actually underestimate the true extent of such changes. Losses of heterozygosity, that is, losses of a maternal or paternal allele in a tumour, are widespread and are often accompanied by a gain of the opposite allele. A tumour could lose the maternal chromosome 8, for example, while duplicating the paternal chromosome 8, leaving the cell with a normal chromosome 8 karyotype but an abnormal chromosome 8 'allelotype'⁴². The 'average' cancer of the colon, breast, pancreas or prostate may lose 25% of its alleles^{42–45} and it is not unusual for a tumour to have lost over half of its alleles (Fig. 4).

Such wholesale changes in the genome of cancer cells might be thought to be due to a true chromosomal instability (CIN), but, as mentioned above, static karyotypes and allelotypes cannot be used to determine rates of genetic alteration. Two types of experiments, however, have shown that these cancers do indeed exhibit a true instability that persists throughout the tumour's lifetime. In one study, fluorescence *in situ* hybridization was used to show that losses or gains of multiple chromosomes occurred 10–100 times more often in aneuploid colorectal cancer cell lines than in normal cells or in diploid cancers of the same histological subtype⁴⁶. In another approach, the rate of loss of heterozygosity of markers surrounding a selectable gene was found to be increased tenfold in a MMR-proficient colorectal cancer cell line compared with a MMR-deficient line⁴⁷. The high rates of chromosome losses and gains seen in these quantitative analyses were consistent with previous observations showing that not only are cancers aneuploid, but that the karyotype of a single tumour is often heterogeneous, reflecting the persistent generation of new chromosomal variations⁵. It has also been shown that the high rate of chromosome gains/losses in aneuploid cancers is not due simply to the ability of aneuploid cells to survive such changes in chromosome number⁴⁶.

Relationship between MIN and CIN. In colorectal and endometrial cancers, there is an inverse relationship between CIN and MIN. Cancers showing MMR deficiency are, in general, diploid and exhibit normal rates of gross chromosomal change, whereas MMR-proficient tumours are usually aneuploid and exhibit increased rates of chromosomal change in the same assays⁴⁶ (Fig. 4). Two other comparisons between MIN and CIN tumours are also informative. First, both types of instability seem to occur early during tumour evolution, but the genetic variations resulting from these instabilities increase with tumour progression. In the case of CIN, aneuploidy can often be observed in small benign

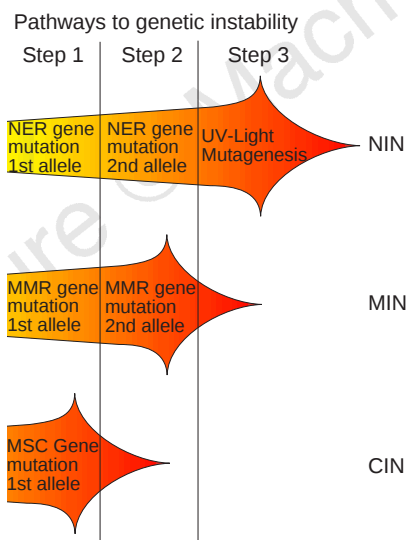


Figure 2 Pathways to genetic instability. Different types of genetic instability require a different number of mutational 'hits' to produce the respective instability phenotype. Top, in a heterozygote with one defective NER allele (step 1), inactivation of the second, normal allele (step 2) does not immediately lead to mutations. Exposure to an environmental agent (ultraviolet light; step 3) is also required to create large numbers of mutations (NER-related instability; NIN). Centre, in contrast, in a heterozygote with one defective MMR allele (step 1), all that is required to begin to develop mutations at a high rate (microsatellite instability, MIN) is the inactivation of the normal allele inherited from the unaffected parent (step 2). Bottom, results of cell fusion and other experiments (see text) indicate that chromosomal instability (CIN) can have a dominant quality. One example of a gene that can be mutated in a dominant-negative manner to cause CIN is *hBUB1*, a component of the mitotic spindle checkpoint (MSC)⁶⁶. Apparently only a single mutational 'hit' of such a gene is required to produce the CIN phenotype.

tumours of the colon or breast, and the deviations from a normal karyotype increase as the tumours enlarge in size and eventually become malignant^{48–50}. Similarly, the microsatellite instability that is the signature of MMR deficiency is seen early during tumorigenesis and increases as the tumours progress in size and disorganization⁵¹. Such increases with progression are expected during clonal expansion, with each new wave of expansion originating in a cell that has divided many times since the last wave.

Second, when MIN and CIN cells are fused, the resulting hybrids (at least in the two cases studied so far; see ref. 46 and C.L., K.W.K. and B.V., unpublished data) exhibit CIN but not MIN. The complementation of MIN by CIN cells is expected, given that expression of the wild-type MMR gene in the hybrid should restore MMR function⁵². The inability of the MIN cells to complement the CIN phenotype, however, shows that the CIN phenotype has a dominant quality, indicating that it may result from gain of function of an expressed protein rather than from gene inactivation. This, in turn, suggests that only a single mutational 'hit' may be required to produce the CIN phenotype (Fig. 2).

Molecular basis of CIN. The molecular basis of CIN in tumours is just beginning to be explored. Because aneuploidy (and, presumably, an underlying CIN) is nearly ubiquitous in cancers⁵ as well as in transformed cells *in vitro*⁵³, it could be argued that CIN results simply from the abnormal structure and growth properties of the cancer cell. However, the fact that CIN generally does not occur in cancer cells with the MIN phenotype, and that tetraploid cells resulting from the fusion of two MIN cell lines remain chromosomally stable, argues strongly against any nonspecific factors⁴⁶.

One potential explanation for CIN invokes p53 mutation. For example, cells in culture often become grossly aneuploid at the same time that p53 is inactivated⁵⁴, and abnormal spindles and arrest in the G2/M phase of the cell cycle occur in p53-deficient cells^{55,56}. It is unlikely, however, that p53 is generally responsible for CIN, as several cancer lines with p53 mutations are diploid and chromosomally stable (and exhibit MIN because of MMR deficiency)^{46,57}. Furthermore, inactivation of p53 by targeted homologous recombination in a non-CIN cancer line has no measurable effect on chromosomal stability (F. Bunz and C.L., unpublished data). Most important, aneuploidy and CIN appear early during tumorigenesis^{48–50}, whereas p53 mutations do not usually occur until much later^{58,59}. These results indicate that although p53 mutations may exacerbate chromosome instability, they are unlikely to be its primary cause.

The critical clue to understanding the molecular basis of MIN in human cancers was provided by studies of unicellular organisms with MMR deficiency (see above). Analogously, the molecular basis for CIN is likely to be unravelled through studies of chromosome

instabilities in yeast and other organisms. However, in contrast to MIN, where only a few genes give rise to the phenotype, a large number of gene alterations can give rise to CIN in yeast^{60–63}. Genes that, when altered, can lead to CIN include those involved in chromosome condensation, sister-chromatid cohesion, kinetochore structure and function and centrosome/microtubule formation and dynamics, as well as 'checkpoint' genes that monitor the proper progression of the cell cycle^{62–65} (Fig. 5).

One type of checkpoint, the spindle checkpoint, ensures that chromatids do not separate until the chromosomes are aligned appropriately along the mitotic spindle. In yeast, disruption of genes encoding spindle-checkpoint proteins can lead to CIN because checkpoint-defective cells can complete mitosis in the presence of a lagging chromosome, resulting in abnormal chromosome segregation^{62–65}. Two lines of evidence have indicated that abnormal spindle-checkpoint genes might be responsible for CIN in human cancers. First, some aneuploid (CIN) lines, but not diploid (MIN) lines, responded aberrantly to spindle-disrupting agents such as nocodazole and colcemid. Instead of undergoing a relatively long-term arrest in metaphase, the CIN lines appeared to exit mitosis prematurely and begin another round of DNA synthesis⁶⁶. The inability to inhibit entry into S phase when mitosis cannot be completed because of spindle damage is the hallmark of a spindle-checkpoint defect^{62–65}. Second, alterations in the expression or sequence of human mitotic-checkpoint genes have been detected in human cancers. For example, decreased expression of *hMAD2* has been seen in breast cancers and might be important in aneuploidy⁶⁷. A small fraction of colorectal cancers have been shown to contain somatic mutations of either *hBUB1* or *hBUBR1* (ref. 66). Mutations in *BUB1* can function in a dominant-negative manner in both mouse and human cells, conferring an abnormal spindle checkpoint when expressed exogenously^{66,68}. In addition, *hMAD1* is a protein that is targeted by the Tax protein of the human T-cell leukaemia virus type 1; this targeting process inactivates the spindle checkpoint in virally induced leukaemias⁶⁹.

Another type of checkpoint, the DNA-damage checkpoint, prevents cells with DNA damage from entering mitosis. Such DNA damage can result from mistakes made by polymerases during normal DNA replication, from exogenous sources (such as ionizing radiation), from endogenous genotoxins (such as reactive oxygen species), or from incomplete repair. Chromosomes containing damaged DNA could segregate inappropriately because sister chromatids are still connected by DNA or DNA–protein links. Such chromosomes are also susceptible to gross structural alterations because of single-stranded gaps or double-stranded-DNA breaks. Studies of yeast have shown that the chromosome instability arising from failed DNA-damage checkpoints is often associated with

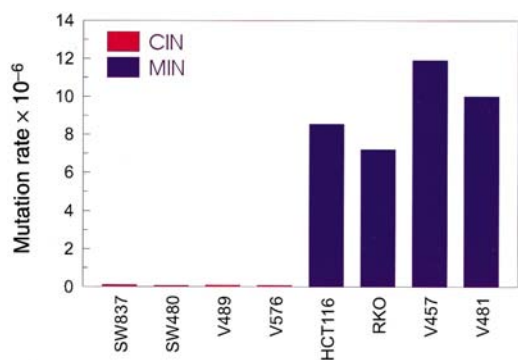


Figure 3 Increased mutation rates at the *HPRT* locus in MIN versus CIN cancer cell lines (ref. 41 and J. Eshleman, M. Veigl, D. Sedwick and S. Markowitz, personal communication). Mutation rates are expressed as 10⁻⁶ mutations per locus per generation.

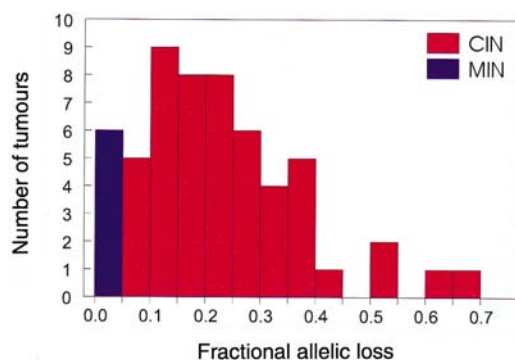


Figure 4 Frequency of allelic losses in MIN versus CIN cancers. Allelic losses were studied by restriction fragment length polymorphisms and Southern blotting. The fractional allelic loss in each tumour is defined as the number of chromosomal arms on which allelic loss was observed divided by the number of chromosomal arms for which allelic markers were informative⁴².

enhanced mitotic recombination as well as with aberrant chromosome segregation, whereas that arising from spindle-checkpoint defects is associated largely with aberrant chromosome segregation⁷⁰. Several genes involved in the DNA-damage checkpoint have been implicated in human tumorigenesis, including *ataxia telangiectasia mutated* (ATM)⁷¹, the ATM-related gene *ATR*⁷², the *BRCA1* and *BRCA2* genes, which interact with the human Rad51 homologue⁷³, and *p53* (ref. 74).

Another potential cause of aneuploidy involves abnormal centrosomes. Multipolar spindles have often been observed in human cancers *in situ* and an abnormal number of centrosomes, detected with specific antibodies, has been observed in breast, lung, prostate, colon, and brain cancers⁷⁵. The abnormal centrosomes in cancers apparently retain the ability to nucleate microtubules *in vitro*. The molecular and genetic bases for the increased number of centrosomes in such cells have not yet been defined. One possibility, however, relates the abnormal centrosomes to a kinase, *aurora2/STK15*, which is involved in centrosome maturation and spindle assembly in *Drosophila*, can affect centrosome number and chromosome segregation when exogenously expressed in mammalian cells, and is highly expressed and occasionally amplified in human cancers^{76,77}. The centrosome-associated kinase PLK1, a homologue of *Drosophila Polo*, has properties similar to those of *aurora2/STK15* and is also highly expressed in a subset of human cancers⁷⁸. Finally, *p53* inactivation can lead to the occurrence of multiple centrosomes in mouse embryonic fibroblasts⁷⁹.

Despite these clues, the molecular basis of CIN remains undefined in most human cancers. The fact that genetic defects of so many genes can lead to CIN, at least in yeast, suggests a heterogeneous basis for CIN in cancers, with many genes each playing a role in a small proportion of the cases (see Supplementary Information). Accordingly, CIN may be so common in cancers precisely

because there are so many genes that, when mutated, can lead to this phenotype and provide the affected cell with the instability required to develop the multiple genetic alterations that lead to malignancy.

Because there are so many genes that may be involved in CIN, it will be essential to use rigorous criteria before concluding that a specific gene is responsible for CIN in a given cancer. In our opinion, these criteria include the following: first, documentation of subtle, intragenic mutations of the gene that alter its function or expression in at least a subset of cases; and second, recapitulation of the CIN phenotype in an immortal diploid line. When the candidate gene is presumed to act in a dominant-negative fashion, recapitulation of the CIN phenotype can be accomplished through exogenous expression of the mutant gene product. When the candidate gene is presumed to act recessively, targeted deletion can be used. In both cases, it is important to do the experiments in immortal, rather than primary, diploid cells to eliminate the possibility that the tested gene is simply immortalizing the cells. Such immortalization, even when spontaneous, may indirectly result in aneuploidy⁵³.

Chromosome translocations

Two patterns of chromosome translocation have been seen in human cancers. The first pattern (complex type) is observed in many solid tumours, but the translocations observed in individual tumours appear nearly random and are usually unrelated to the translocations observed in other tumours of the same histological subtype^{5,80}. Marker chromosomes, containing complicated rearrangements of several chromosomes, are seen frequently in such tumours. Large portions of chromosomal arms are often deleted during the recombinations that lead to the translocations; these deletions are seen as losses of heterozygosity at the molecular level. Losses of heterozygosity can also be generated by mitotic recombinations, which are invisible at the karyotypic level but occur very commonly in most solid tumours (Fig. 4).

The complex type of translocation can result in gains and losses of chromosomal material, much like CIN, as well as in the generation of new gene products, and is likely to be an important component of many common tumours. However, the rate of such translocations in cells cannot be reliably measured, so it is not clear how much more frequently translocations occur in neoplastic cells compared with normal cells. Even the rate of mitotic recombination, which is easy to measure in yeast through analysis of the progeny, cannot be assessed simply in mammalian cells. The molecular basis for the translocations in cancers is not known. However, one attractive possibility is that they arise in cells that enter mitosis before recombination-promoting double-strand breaks (DSBs) are repaired^{64,65}. Good candidates for mediating such 'translocation instability' in human cancers are, therefore, genes such as *ATM*, *ATR*, *BRCA1*, *BRCA2*, *p53* and other components involved in DSB repair or DNA-damage checkpoints (see above). To determine whether these or other candidate genes are actually responsible for translocation instability, we will need convenient assays with which to measure the rate of translocation in somatic cells and more detailed characterizations of the recombination joints. It is not yet known, for example, whether the joints in complex translocations result from homologous or non-homologous recombination. Results from some studies indicate that inactivation of *p53* may increase the rate of homologous recombination of experimental substrates^{81,82}, thus providing yet another link between *p53* and increased genetic instability at late stages of tumorigenesis.

A second pattern of translocation (simple type) is characterized by distinctive rearrangements of chromosomal segments in specific neoplastic diseases. These occur commonly in leukaemias and lymphomas and in a subset of sarcomas and other rare tumours⁸³. The simple type of translocation reproducibly involves the same chromosomal segments, so much so that their karyotypic presence can be used to classify neoplasms and predict therapeutic responses⁸³. Molecular analyses have determined the breakpoints

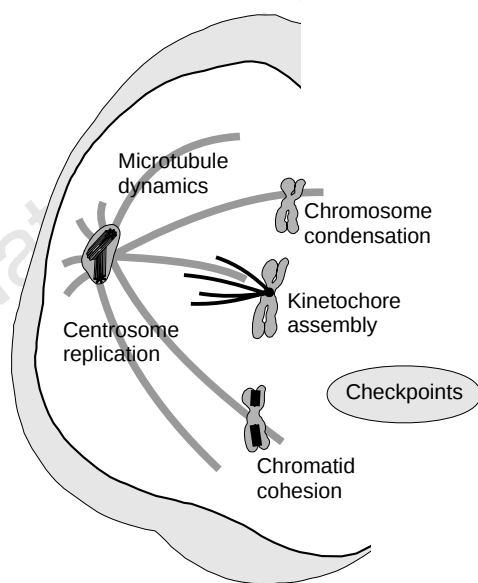


Figure 5 Cellular processes involved in replication and segregation of chromosomes during mitosis. Processes involved include chromosome condensation, cohesion of sister chromatids, and centrosome/microtubule formation and dynamics. Checkpoints that are required in chromosome replication and segregation include the mitotic spindle checkpoint, which ensures that chromosomes are aligned correctly before anaphase, and the DNA-damage checkpoint, which prevents cells with DNA damage from entering prophase. Aberrations in these processes and checkpoints could give rise to the CIN phenotype. Numerous examples of specific genes that could lead to CIN when mutated are listed in Supplementary Information.

in many of the translocations and have shown that these usually occur within the same, relatively small segments of DNA. The rearrangements generally lead to the activation of an oncogene through its positioning near a strong promoter or its fusion with another gene. These specific translocations appear to be essential for the development or progression of the neoplasms in which they occur. However, there is no evidence that they are the result of a specific genetic instability that promotes their occurrence at higher frequency than in normal cells. In fact, some of these translocations represent an aberration of the normal recombination processes mediated by RAG proteins that lead to immunoglobulin or T-cell-receptor gene rearrangements, and similar translocations occur in normal cells of lymphoid origin^{84,85}.

Gene amplifications

Amplifications of oncogenes occur in a subset of late-stage cancers of many organs, and amplification of genes involved in metabolism or inactivation of drugs represents a common way for cultured cells to acquire resistance to chemotherapeutic agents⁹. Quantitative measurements have indicated that amplifications may occur at a higher rate in cancer cells than in normal cells⁸⁶. The mechanisms through which amplifications are generated are largely unknown. Further progress in this area might be made if an analogous process could be demonstrated in yeast, an organism that has provided key insights into other forms of genetic change. One potential player in the process is p53, as amplifications occur more easily when p53 is inactivated in mammalian cells^{87,88}. This is not the complete story, however, as amplification can also occur in cancer cells with wild-type p53 genes⁸⁷. One theory is that it is not the rate at which amplification occurs that distinguishes normal cells from cancer cells, but rather the ability of cancer cells exhibiting gene amplification to survive. In normal cells, the presence of an amplicon (the genomic segment that is amplified) may signal that DNA damage has occurred, thus setting off p53-dependent apoptosis⁸⁹. In the absence of normal p53, cells with initial amplifications of a locus may survive and accumulate additional amplicons during further rounds of cellular division. This model would distinguish 'amplification instability' from CIN, in which the actual rate of chromosome-number change, rather than the ability of cells to survive such changes, is altered⁴⁶.

Gene amplification is an important process in human cancers, as it is clearly associated with tumour progression⁹, has prognostic significance¹⁰ and has even provided a target for therapeutics in the case of amplification of HER2/neu in breast cancers⁹⁰. However, it is important to distinguish gene amplification from the other forms of genetic change discussed above. NIN, MIN and CIN occur early in the neoplastic process, affect multiple genes in each cell, and are likely to be essential for development of the sequential mutations that drive the neoplastic process. Although amplifications can significantly affect tumour biology, they affect only a single or a few genes in each cell and, in general, occur late in tumorigenesis.

Conclusions

The data reviewed above lead us to conclude that nearly all solid tumours are genetically unstable. In most cases, the instability is seen at the chromosomal level, with frequent gains and losses of whole chromosomes (CIN). In a few cases, the instability is at the nucleotide level (NIN or MIN) and is the result of faulty DNA repair. Translocations and amplifications add to the chromosomal abnormalities, and may reflect additional mechanisms for generating instability that occur as tumours progress. Instability is the engine of both tumour progression and tumour heterogeneity, guaranteeing that no two tumours are exactly alike and that no single tumour is composed of genetically identical cells. This heterogeneity undermines otherwise sound therapeutic strategies⁹¹.

But there is another side to genetic instability, one that may work to the patient's benefit rather than to the tumour's. Much effort is

currently being expended to target, during treatment, the mutated oncogenes and tumour-suppressor genes that control neoplastic cell growth directly. The genetic instabilities that underlie cancer may provide equally valid therapeutic targets. Although such instabilities are not directly responsible for the cancer's abnormal growth, they are likely to be genetically based and, therefore, permanent components of the cancer cell that unambiguously distinguish it from all normal cells. Because instabilities reflect defects in cellular processes that maintain the integrity of the genome, they can be expected to generate sensitivities to particular chemical agents. For example, cells with defects in nucleotide-excision repair are sensitive to ultraviolet light¹⁶, and cells with defective *BRCA* genes are sensitive to ionizing radiation (reviewed in ref. 92). It can be expected that drugs will be found to which mismatch-repair-deficient cells are particularly sensitive and that some cells with defective mitotic checkpoints will be more sensitive to microtubule-disrupting agents (such as paclitaxel) than normal cells. In fact, one can argue persuasively that all chemotherapeutic compounds used at present are more toxic to cancer cells than to normal cells only and specifically because of the defective checkpoints that occur in the former cells^{2,93}. This line of reasoning suggests that, although instability may be essential for neoplasia to develop, it may also prove to be its Achilles' heel when the tumour is attacked by the right agents. Further research to define the molecular and physiologic bases of instability may, therefore, yield entirely new approaches to treating common forms of cancer. □

C. Lengauer, K. W. Kinzler and B. Vogelstein are at the Johns Hopkins Oncology Center and B. Vogelstein is at the Howard Hughes Medical Institute, Baltimore, Maryland 21231, USA.

- Loeb, L. A. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.* **51**, 3075–3079 (1991).
- Hartwell, L. Defects in a cell cycle checkpoint may be responsible for the genomic instability of cancer cells. *Cell* **71**, 543–546 (1992).
- Tomlinson, I. P., Novelli, M. R. & Bodmer, W. F. The mutation rate and cancer. *Proc. Natl Acad. Sci. USA* **93**, 14800–14803 (1996).
- Almoguer, C. et al. Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. *Cell* **53**, 549–554 (1988).
- Mitelman, F., Johansson, B. & Mertens, F. *Catalog of Chromosome Aberrations in Cancer* Vol. 2 (Wiley-Liss, New York, 1994).
- Wang, S. I. et al. Somatic mutations of PTEN in glioblastoma multiforme. *Cancer Res.* **57**, 4183–4186 (1997).
- Zhuang, Z. et al. Trisomy 7-harboring non-random duplication of the mutant MET allele in hereditary papillary renal carcinomas. *Nature Genet.* **20**, 66–69 (1998).
- Nowell, P. C. Genetic alterations in leukemias and lymphomas: impressive progress and continuing complexity. *Cancer Genet. Cytogenet.* **94**, 13–19 (1997).
- Brodeur, G. M. & Hogarty, M. D. in *The Genetic Basis of Human Cancer* 1st edn Vol. 1 (eds Kinzler, K. W. & Vogelstein, B.) 161–179 (McGraw-Hill, New York, 1998).
- Seeger, R. C. et al. Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastomas. *N. Engl. J. Med.* **313**, 1111–1116 (1985).
- Kauffmann-Zeh, A. et al. Suppression of c-Myc-induced apoptosis by Ras signalling through PI(3)K and PKB. *Nature* **385**, 544–548 (1997).
- Kunkel, T. A. DNA mismatch repair. The intricacies of eukaryotic spell-checking. *Curr. Biol.* **5**, 1091–1094 (1995).
- Sia, E. A., Jinks-Robertson, S. & Petes, T. D. Genetic control of microsatellite stability. *Mutat. Res.* **383**, 61–70 (1997).
- Cleaver, J. E. Defective repair replication of DNA in xeroderma pigmentosum. *Nature* **218**, 652–656 (1968).
- De Weerd-Kastelein, E. A., Keijzer, W. & Bootsma, D. Genetic heterogeneity of xeroderma pigmentosum demonstrated by somatic cell hybridization. *Nat. New Biol.* **238**, 80–83 (1972).
- Bootsma, D., Kraemer, K. H., Cleaver, J. E. & Hoeijmakers, J. H. J. in *The Genetic Basis of Human Cancer* (eds Kinzler, K. W. & Vogelstein, B.) 245–274 (McGraw-Hill, New York, 1998).
- Wood, R. D. DNA repair in eukaryotes. *Annu. Rev. Biochem.* **65**, 135–167 (1996).
- Cairns, J. The origin of human cancers. *Nature* **289**, 353–357 (1981).
- Feinberg, A. P. & Coffey, D. S. Organ site specificity for cancer in chromosomal instability disorders. *Cancer Res.* **42**, 3252–3254 (1982).
- Kraemer, K. H., Lee, M. M. & Scotto, J. DNA repair protects against cutaneous and internal neoplasia: evidence from xeroderma pigmentosum. *Carcinogenesis* **5**, 511–514 (1984).
- Ames, B. N. & Gold, L. S. Endogenous mutagens and the causes of aging and cancer. *Mutat. Res.* **250**, 3–16 (1991).
- Peinado, M. A., Malkhosyan, S., Velazquez, A. & Perucho, M. Isolation and characterization of allelic losses and gains in colorectal tumors by arbitrarily primed polymerase chain reaction. *Proc. Natl Acad. Sci. USA* **89**, 10065–10069 (1992).
- Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. & Perucho, M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363**, 558–561 (1993).
- Thibodeau, S. N., Bren, G. & Schaid, D. Microsatellite instability in cancer of the proximal colon. *Science* **260**, 816–819 (1993).
- Aaltonen, L. A. et al. Clues to the pathogenesis of familial colorectal cancer. *Science* **260**, 812–816 (1993).
- Peltomäki, P. et al. Genetic mapping of a locus predisposing to human colorectal cancer. *Science* **260**, 810–812 (1993).

27. Lindblom, A., Tannergard, P., Werelius, B. & Nordenskjold, M. Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nature Genet.* **5**, 279–282 (1993).
28. Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274–276 (1993).
29. Fishel, R. *et al.* The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75**, 1027–1038 (1993).
30. Leach, F. S. *et al.* Mutations of a mutL homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**, 1215–1225 (1993).
31. Bronner, C. E. *et al.* Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* **368**, 258–261 (1994).
32. Papadopoulos, N. *et al.* Mutation of a mutL homolog in hereditary colon cancer. *Science* **263**, 1625–1629 (1994).
33. Parsons, R. *et al.* Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* **75**, 1227–1236 (1993).
34. Umar, A. *et al.* Defective mismatch repair in extracts of colorectal and endometrial cancer cell lines exhibiting microsatellite instability. *J. Biol. Chem.* **269**, 14367–14370 (1994).
35. Peltomäki, P. & de la Chapelle, A. Mutations predisposing to hereditary nonpolyposis colorectal cancer. *Adv. Cancer Res.* **71**, 93–119 (1997).
36. Modrich, P. Mismatch repair, genetic stability and tumour avoidance. *Phil. Trans. R. Soc. Lond. B* **347**, 89–95 (1995).
37. Kolodner, R. Biochemistry and genetics of eukaryotic mismatch repair. *Genes Dev.* **10**, 1433–1442 (1996).
38. Perucho, M. Cancer of the microsatellite mutator phenotype. *Biol. Chem.* **377**, 675–684 (1996).
39. Dams, E., Van de Kelft, E. J., Martin, J. J., Verlooy, J. & Willems, P. J. Instability of microsatellites in human gliomas. *Cancer Res.* **55**, 1547–1549 (1995).
40. Bhattacharya, N. P., Skandalis, A., Ganesh, A., Groden, J. & Meuth, M. Mutator phenotypes in human colorectal carcinoma cell lines. *Proc. Natl Acad. Sci. USA* **91**, 6319–6323 (1994).
41. Eshleman, J. R. *et al.* Increased mutation rate at the hprt locus accompanies microsatellite instability in colon cancer. *Oncogene* **10**, 33–37 (1995).
42. Vogelstein, B. *et al.* Allelotype of colorectal carcinomas. *Science* **244**, 207–211 (1989).
43. Boige, V. *et al.* Concerted nonsynthetic allelic losses in hyperploid hepatocellular carcinoma as determined by a high-resolution allelotype. *Cancer Res.* **57**, 1986–1990 (1997).
44. Seymour, A. B. *et al.* Allelotype of pancreatic adenocarcinoma. *Cancer Res.* **54**, 2761–2764 (1994).
45. Radford, D. M. *et al.* Allelotyping of ductal carcinoma *in situ* of the breast: deletion of loci on 8p, 13q, 16q, 17p and 17q. *Cancer Res.* **55**, 3399–3405 (1995).
46. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386**, 623–627 (1997).
47. Phear, G., Bhattacharya, N. P. & Meuth, M. Loss of heterozygosity and base substitution at the APRT locus in mismatch-repair-proficient and -deficient colorectal carcinoma cell lines. *Mol. Cell. Biol.* **16**, 6516–6523 (1996).
48. Bardi, G. *et al.* Cytogenetic comparisons of synchronous carcinomas and polyps in patients with colorectal cancer. *Br. J. Cancer.* **76**, 765–769 (1997).
49. Silverstein, M. J. *Ductal Carcinoma in situ of the Breast* (Williams & Wilkins, Baltimore, 1997).
50. Bomme, L. *et al.* Cytogenetic analysis of colorectal adenomas: karyotypic comparisons of synchronous tumors. *Cancer Genet. Cytogenet.* **106**, 66–71 (1998).
51. Aaltonen, L. A. *et al.* Replication errors in benign and malignant tumors from hereditary non-polyposis colorectal cancer patients. *Cancer Res.* **54**, 1645–1648 (1994).
52. Koi, M. *et al.* Human chromosome 3 corrects mismatch repair deficiency and microsatellite instability and reduces N-methyl-N'-nitro-N-nitrosoguanidine tolerance in colon tumor cells with homozygous hMLH1 mutation. *Cancer Res.* **54**, 4308–4312 (1994).
53. Li, R. *et al.* Aneuploidy correlated 100% with chemical transformation of Chinese hamster cells. *Proc. Natl Acad. Sci. USA* **94**, 14506–14511 (1997).
54. Harvey, M. *et al.* *In vitro* growth characteristics of embryo fibroblasts isolated from p53-deficient mice. *Oncogene* **8**, 2457–2467 (1993).
55. Cross, S. M. *et al.* A p53-dependent mouse spindle checkpoint. *Science* **267**, 1353–1356 (1995).
56. Lanni, J. S. & Jacks, T. Characterization of the p53-dependent postmitotic checkpoint following spindle disruption. *Mol. Cell. Biol.* **18**, 1055–1064 (1998).
57. Eshleman, J. R. *et al.* Chromosome number and structure both are markedly stable in RER colorectal cancers and are not destabilized by mutation of p53. *Oncogene* **17**, 719–725 (1998).
58. Baker, S. J. *et al.* p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res.* **50**, 7717–7722 (1990).
59. Auer, G. U., Heselmeyer, K. M., Steinbeck, R. G., Munck-Wikland, E. & Zetterberg, A. D. The relationship between aneuploidy and p53 overexpression during genesis of colorectal adenocarcinoma. *Virchows Arch.* **424**, 343–347 (1994).
60. Hoyt, M. A., Stearns, T. & Botstein, D. Chromosome instability mutants of *Saccharomyces cerevisiae* that are defective in microtubule-mediated processes. *Mol. Cell. Biol.* **10**, 223–234 (1990).
61. Spencer, F., Gerring, S. L., Connelly, C. & Hieter, P. Mitotic chromosome transmission fidelity mutants in *Saccharomyces cerevisiae*. *Genetics* **124**, 237–249 (1990).
62. Murray, A. W. The genetics of cell cycle checkpoints. *Curr. Opin. Genet. Dev.* **5**, 5–11 (1995).
63. Nasmyth, K. At the heart of the budding yeast cell cycle. *Trends Genet.* **12**, 405–412 (1996).
64. Elledge, S. J. Cell cycle checkpoints: preventing an identity crisis. *Science* **274**, 1664–1672 (1996).
65. Paulovich, A. G., Toczyski, D. P. & Hartwell, L. H. When checkpoints fail. *Cell* **88**, 315–321 (1997).
66. Cahill, D. P. *et al.* Mutations of mitotic checkpoint genes in human cancers. *Nature* **392**, 300–303 (1998).
67. Li, Y. & Benezra, R. Identification of a human mitotic checkpoint gene: hSMAD2. *Science* **274**, 246–248 (1996).
68. Taylor, S. S. & McKeon, F. Kenetochore localization of murine Bub1 is required for normal mitotic timing and checkpoint response to spindle damage. *Cell* **89**, 727–735 (1997).
69. Jin, D. Y., Spencer, F. & Jeang, K. T. Human T cell leukemia virus type 1 oncoprotein Tax targets the human mitotic checkpoint protein MAD1. *Cell* **93**, 81–91 (1998).
70. Hartwell, L. H. & Smith, D. Altered fidelity of mitotic chromosome transmission in cell cycle mutants of *S. cerevisiae*. *Genetics* **110**, 381–395 (1985).
71. Rotman, G. & Shiloh, Y. ATM: from gene to function. *Hum. Mol. Genet.* **7**, 1555–1563 (1998).
72. Smith, L. *et al.* Duplication of ATR inhibits MyoD, induces aneuploidy and eliminates radiation-induced G1 arrest. *Nature Genet.* **19**, 39–46 (1998).
73. Zhang, H., Tomblin, G. & Weber, B. L. BRCA1, BRCA2, and DNA damage response: collision or collusion? *Cell* **92**, 433–436 (1998).
74. Lane, D. Awakening angels. *Nature* **394**, 616–617 (1998).
75. Doxsey, S. The centrosome—a tiny organelle with big potential. *Nature Genet.* **20**, 104–106 (1998).
76. Zhou, H. *et al.* Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nature Genet.* **20**, 189–193 (1998).
77. Bischoff, J. R. *et al.* A homologue of *Drosophila* aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J.* **17**, 3052–3065 (1998).
78. Wolf, G. *et al.* Prognostic significance of polo-like kinase (PLK) expression in non-small cell lung cancer. *Oncogene* **14**, 543–549 (1997).
79. Fukasawa, K., Choi, T., Kuriyama, R., Rulong, S. & Vande Woude, G. F. Abnormal centrosome amplification in the absence of p53. *Science* **271**, 1744–1747 (1996).
80. Johansson, B., Mertens, F. & Mitelman, F. Primary vs. secondary neoplasia-associated chromosomal abnormalities—balanced rearrangements vs. genomic imbalances? *Genes Chromosomes Cancer* **16**, 155–163 (1996).
81. Sturzbecher, H. W., Donzelmann, B., Henning, W., Knippschild, U. & Buchhop, S. p53 is linked directly to homologous recombination processes via RAD51/RecA protein interaction. *EMBO J.* **15**, 1992–2002 (1996).
82. Meekel, K. L. *et al.* Inactivation of p53 results in high rates of homologous recombination. *Oncogene* **14**, 1847–1857 (1997).
83. Le Beau, M. M. & Rowley, J. D. Chromosomal abnormalities in leukemia and lymphoma: clinical and biological significance. *Adv. Hum. Genet.* **15**, 1–54 (1986).
84. Tycko, B., Palmer, J. D. & Sklar, J. T cell receptor gene trans-rearrangements: chimeric gamma-delta genes in normal lymphoid tissues. *Science* **245**, 1242–1246 (1989).
85. Hiom, K., Melek, M. & Gellert, M. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* **94**, 463–470 (1998).
86. Tlsty, T. D., Margolin, B. H. & Lum, K. Differences in the rates of gene amplification in nontumorigenic and tumorigenic cell lines as measured by Luria-Delbruck fluctuation analysis. *Proc. Natl Acad. Sci. USA* **86**, 9441–9445 (1989).
87. Livingston, L. R. *et al.* Altered cell cycle arrest and gene amplification potential accompany loss of wild-type p53. *Cell* **70**, 923–935 (1992).
88. Yin, Y., Tainsky, M. A., Bischoff, F. Z., Strong, L. C. & Wahl, G. M. Wild-type p53 restores cell cycle control and inhibits gene amplification in cells with mutant p53 alleles. *Cell* **70**, 937–948 (1992).
89. Oren, M. Relationship of p53 to the control of apoptotic cell death. *Semin. Cancer Biol.* **5**, 221–227 (1994).
90. Pegram, M. D. *et al.* Phase II study of receptor-enhanced chemosensitivity using recombinant humanized anti-p185HER2/neu monoclonal antibody plus cisplatin in patients with HER2/neu-overexpressing metastatic breast cancer refractory to chemotherapy treatment. *J. Clin. Oncol.* **16**, 2659–2671 (1998).
91. Owens, A. H., Coffey, D. S. & Baylin, S. B. *Tumor Cell Heterogeneity* (Academic, New York, 1982).
92. Abbott, D. W., Freeman, M. L. & Holt, J. T. Double-strand break repair deficiency and radiation sensitivity in BRCA2 mutant cancer cells. *J. Natl Cancer Inst.* **90**, 978–985 (1998).
93. Waldman, T., Lengauer, C., Kinzler, K. W. & Vogelstein, B. Uncoupling of S phase and mitosis induced by anticancer agents in cells lacking p21. *Nature* **381**, 713–716 (1996).
94. Williams, C. *et al.* Clones of normal keratinocytes and a variety of simultaneously present epidermal neoplastic lesions contain a multitude of p53 gene mutations in a xeroderma pigmentosum patient. *Cancer Res.* **58**, 2449–2455 (1998).
95. Markowitz, S. *et al.* Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* **268**, 1336–1338 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank S. Markowitz and A. Weith for sharing unpublished data, and our colleagues in the Molecular Genetics Laboratory for reviewing the manuscript. This work was supported by the Clayton Fund and grants from the National Cancer Institute.

Correspondence and requests for materials should be addressed to C.L. (e-mail: lengauer@jhmi.edu).

Coexisting conical bipolar and equatorial outflows from a high-mass protostar

L. J. Greenhill*, C. R. Gwinn†, C. Schwartz‡§, J. M. Moran* & P. J. Diamond||

* Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA

† University of California, Department of Physics, Broida Hall, Santa Barbara, California 93106, USA

‡ Maria Mitchell Observatory, 3 Vestal Street, Nantucket, Massachusetts 02554, USA

|| National Radio Astronomy Observatory, PO Box O, Socorro, New Mexico 87801, USA

The BN/KL region in the Orion molecular cloud¹ is an archetype for the study of the formation of stars much more massive than the Sun². This region contains luminous young stars and protostars but, like most star-forming regions, is difficult to study in detail because of the obscuring effects of dust and gas. Our basic expectations are shaped to some extent by the present theoretical picture of star formation, the cornerstone of which is that protostars accrete gas from rotating equatorial disks and shed angular momentum by ejecting gas in bipolar outflows. The main source of the outflow in the BN/KL region^{3–5} may be an object known as radio source I (ref. 6), which is commonly believed to be surrounded by a rotating disk of molecular material^{7–9}. Here we report high-resolution observations of silicon monoxide (SiO) and water maser emission from the gas surrounding source I. We show that within 60 AU of the source (about the size of the Solar System), the region is dominated by a conical bipolar outflow, rather than the expected disk. A slower outflow, close to the equatorial plane of the protostellar system, extends to radii of 1,000 AU.

The BN/KL region radiates an estimated $(5\text{--}8) \times 10^4 \text{ erg s}^{-1}$ (refs 4, 10), and is a rich source of molecular line emission^{2,11} embedded within the Orion molecular cloud (OMC-1). Two molecular outflows originate in BN/KL relatively close to source I: a low-velocity (18 km s^{-1}) outflow to the northeast–southwest^{11,12} and an orthogonal high-velocity ($30\text{--}100 \text{ km s}^{-1}$) outflow^{11–14}, both of which extend over $\sim 1'$ (0.14 pc). Source I has a broad-band radio spectrum consistent with an H II region $\sim 60 \text{ AU}$ in diameter¹⁵, though it has not yet been specifically detected at infrared wavelengths. Masers are often associated with massive young stars, and many are broadly distributed across BN/KL. (Masers are the microwave equivalent of lasers.) However, a distinct population of SiO and H₂O masers is clustered around source I^{2,3}. To understand better the physical nature of source I and its relationship to the outflows in BN/KL, we have mapped these masers with high spectral and angular resolution. We resolve the angular structure of the SiO emitting region in detail at each observed line-of-sight velocity, in contrast to previous observations with connected-element radio interferometers^{5,7,8,16}.

Maser sources are valuable tracers in molecule-rich regions because they comprise many compact, high-surface-brightness, spatially distinct points of radiation (spots), with various line-of-sight velocities. The spots are test ‘particles’ tracing the three-dimensional velocity field of the material in which they are embedded. Maser amplification requires relatively small gradients in line-of-sight velocity to achieve long gain-path lengths. To pump the maser levels by collisions, the gas temperature must also be at least a few hundred degrees, and the molecular hydrogen density

orders of magnitude greater than that sampled by thermal molecular emission, that is, $n_{\text{H}_2} \geq 10^8 \text{ cm}^{-3}$. Maser emission is a selective tracer, but in star-forming regions it can in principle occur in either outflows or disks¹⁷.

We observed emission from the $\nu = 1, J = 1\text{--}0$ transition of SiO (43.12208 GHz rest frequency) towards the BN/KL region on 9 July 1995 with the Very Long Baseline Array (VLBA) of the NRAO. The data were processed with the VLBA correlator (which provided spectral channels 0.22 km s^{-1} wide), and reduced with standard techniques¹⁸. Within 30 km s^{-1} of the 5 km s^{-1} velocity of the region near source I², we detected 1,212 maser spots with flux densities between 0.09 and 300 Jy .

Spectroscopic images show that the SiO maser emission is confined to four regions that can be thought of as the arms of an ‘X’, subtending $\sim 0.25''$ (Fig. 1a). We propose that the ‘X’ marks the intersection of two outflow cones with the plane of the sky and separates the low-velocity (northeast–southwest) and high-velocity (northwest–southeast) outflows in BN/KL. Source I was not detected by us because its surface brightness is too low. However, to estimate the position of source I on our maps, we compared the flux-weighted emission centroid for each VLBA spectral channel to the earlier, lower-resolution ($0.25''$) map of Menten and Reid⁵; these workers measured the position of source I relative to the maser emission centroids with $0.01''$ uncertainty, using the Very Large Array (VLA). The relative positions of the dominant clumps of masers in the two maps agree to $0.01''$. Source I lies at the centre of the ‘X’, within the measurement errors (Fig. 1a). The innermost SiO maser features are separated by about the diameter (60 AU) of the source I H II region.

We have added to our study archived H₂O maser data collected on 25 August 1983 with the VLA (Fig. 1b), as well as published maps¹⁹ of the H₂O maser distribution on 5 August 1991. The distribution of the H₂O masers is elongated in the direction of the low-velocity outflow in BN/KL. Over eight years, the length of the distribution grew by $\sim 0.2''$. We measured proper motions for 19 individual maser features that persisted over the eight-year period (Fig. 1b). The motions are primarily parallel to the axis of the low-velocity outflow, and the average transverse motion corresponds to $\sim 20 \text{ km s}^{-1}$, approximately the speed of the large-scale low-velocity outflow in BN/KL¹².

The previous widely accepted model, in which the SiO masers lie in a rotating, expanding and inclined disk^{7–9}, does not fit our high-resolution VLBA maps. One attraction of this previous model was that the disk could feasibly extend to larger radii, explaining the low-velocity BN/KL outflow. In such a disk, the SiO emission is expected to trace dynamic surfaces where the derivative of the line-of-sight velocity is small, so that the maser gain is high. For most rotation laws, each arm of the ‘X’ would be expected to show a velocity gradient along its length; but three of the four do not. Moreover, the distribution of SiO masers that we observe shows approximate mirror symmetry in position and velocity about a northwest–southeast axis, which constrains the magnitude of any rotation to be less than a few kilometres per second. If the disk is gravitationally bound to the protostar, then this rotation speed, at $25\text{--}60 \text{ AU}$ radius, constrains the protostellar mass to be less than 1 solar mass. If radial forcing (for example, a wind) enlarges the disk from an initially bound state (with about 1 AU radius), then conservation of angular momentum requires a stellar mass less than a few solar masses. These relatively small stellar masses are inconsistent with the presence of an H II region.

We propose that the SiO maser emission outlines the limbs of a bipolar, conical, high-velocity flow $25\text{--}60 \text{ AU}$ from the protostar (that is, the ‘X’), and that the H₂O masers lie in a low-velocity flow in an equatorial band bisecting the high-velocity flow (Fig. 1c). The position angles of the cones are -40° and 150° , with opening angles of 60° and 80° , respectively. For an outflow axis close to the plane of the sky and a finite cone-wall thickness, the longest maser gain

§ Present address: Colby College, Waterville, Maine 04901, USA.

paths lie along the limbs of the cones. The difference in mean Doppler velocity between the red- and blue-shifted SiO masers ($10\text{--}15\text{ km s}^{-1}$) may be explained if the cones are tipped out of the plane of the sky. The known distribution of H_2 emission in BN/KL²⁰ shows two lobes, with line-of-sight velocities Doppler-shifted by more than 30 km s^{-1} , that subtend the openings of the cones. The north-

west lobe is equally prominent in red- and blue-shifted emission, which suggests that the inclination is small. (Blue-shifted emission from the southeast lobe appears to be blocked by a dense foreground cloud^{15,21}.)

The dense gas in which the SiO masers lie is probably not gravitationally bound to the protostar, since there is no detectable

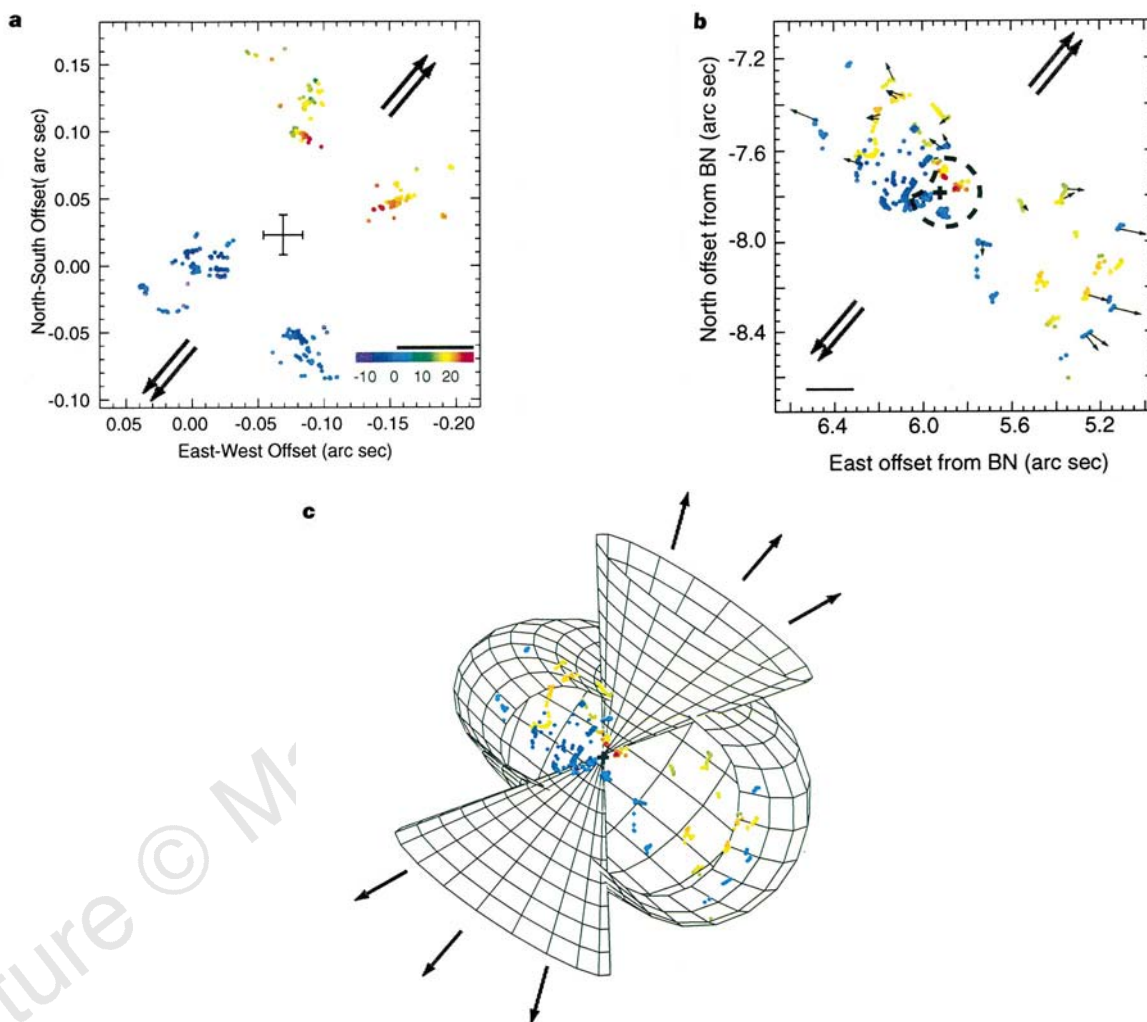


Figure 1 Maser spots tracing outflowing gas in the BN/KL region of Orion. **a**, The $v = 1, J = 1-0$ SiO maser distribution in the BN/KL region, obtained with the VLBA. The VLBA resolves the angular structure of the distribution much more fully than do the connected-element interferometers used in other studies. The cross represents the centroid and position uncertainty for continuum source I with respect to the masers. The double arrows indicate the direction of the proposed high-velocity outflow from source I, and the approximate direction of the high-velocity outflow in BN/KL. Symbol colour indicates Doppler velocity (in km s^{-1}) with respect to the local standard of rest, as indicated by the coloured bar. The black bar indicates 25 AU, for an assumed distance of 450 pc. The interferometer beam is 0.40×0.16 mas (0.18×0.07 AU) and the relative position uncertainties for the masers are smaller than the plotted symbols. We referenced the data to the strong emission at -3.4 km s^{-1} (at the origin) to stabilize the VLBA against atmospheric fluctuations along the line of sight. Deconvolution of the instrumental response required special attention because most images contained complex structure, and because of the proximity of Orion to the celestial equator. From our VLBA data, we estimate that the origin of the map is right ascension (RA) 5 h 35 min 14.5101 (± 0.0003) s, declination (dec.) $-05^\circ 22' 30.63(\pm 0.05)''$ (equinox 2000). **b**, The distribution of H_2O masers in BN/KL within $2''$ of source I, observed with the VLA in 1983, at $0.1''$ resolution. The map of the SiO maser emission is shown within the dashed contour. The cross marks the position of source I. The registration uncertainty between the H_2O and SiO distributions is $\sim 0.05''$, limited by uncertainty in the astrometric position of each with respect to the extragalactic radio reference frame. Arrows attached to the H_2O maser spots

show the estimated relative proper motions over 8 years. The mean motion has been subtracted to account for motion of the reference emission. The double arrows indicate the direction of the proposed high-velocity outflow from source I. The scale bar indicates 100 AU; an arrow as long as the bar corresponds to a motion of $\sim 50\text{ km s}^{-1}$. The 1983 VLA observations covered a total velocity range of -12.4 to 28.1 km s^{-1} , except for velocities close to a $4 \times 10^5\text{ Jy}$ H_2O maser flare³² with a velocity of 8 km s^{-1} and a position of RA 5 h 35 min 14.121 (± 0.003) s, dec. $-05^\circ 22' 36.27(\pm 0.05)''$ (equinox 2000). The maser data were self-calibrated and relative positions for the maser spots measured. Absolute positions were obtained from low-dynamic-range images obtained at the same time by a wide-band antenna subarray that observed the maser flare and a phase calibrator, 0529 + 0.75. **c**, A proposed model for the region surrounding source I, showing the high-velocity and low-velocity outflows schematically as cones and a non-axisymmetric torus, respectively. The front surface of the torus is cut away for clarity, and the SiO and H_2O maser distributions are superposed. The arrows indicate schematically the high-velocity outflow; the arrows at position angles of -60° and 120° also correspond approximately to the direction of the magnetic field in OMC-1²⁸. A toroidal geometry is shown but is not the only possibility. The H_2O masers lie close to the surface of the torus, probably in outward-propagating shocks. Overall, the zone in which the H_2O masers lie may not move significantly over time, as individual shocks move outwards and diminish over time, while new ones are born closer to the protostar. Otherwise, the dynamical age of the equatorial outflow is of the order of 100 years.

rotation of the gas. The SiO maser emission probably arises at a steady-state shock interface, perhaps between a faster and a slower component of the stellar outflow. Such shocks serve as efficient mechanisms that collisionally pump SiO masers and reduce depletion of gas-phase SiO onto dust grains. To quantify the lifetime of the visible structure, we compared the present SiO maser maps with a partially complete VLBI map obtained in 1985²². The loci of red-shifted masers have remained fixed with respect to source I to $\sim 0.01''$ over 10.3 years, which corresponds to a proper motion of less than 4.5 AU or 2 km s^{-1} . We speculate that collimation of the bipolar outflow begins at radii less than 25 AU, perhaps inside the H II region, because source I lies close to the apices of the outflow cones.

We fitted the H₂O maser positions and Doppler velocities to a constant-velocity outflow model centred on source I, where the H₂O masers are constrained to lie on a toroidal surface, as suggested qualitatively by the curvature traced by several H₂O features (Fig. 1c). This surface may be where large-diameter dust grains begin to form and to be accelerated by radiation pressure. (Dust condensation and H₂O maser excitation together require temperatures of 400–1,000 K. If this material is in thermal equilibrium, then the luminosity of source I would be $\gg 10^4$ solar luminosities.) The torus has the following parameters: mean outer radius 360 AU, thickness 270–320 AU, outflow velocity 13 km s^{-1} , and rotation less than a few kilometres per second. The torus is tipped down $\sim 10^\circ$ with respect to the line of sight, as evidenced by the southeasterly position bias of the distribution of blue-shifted H₂O masers. (The masers lie close to the equator.) We assume that the axes of the high-velocity outflow and the torus are aligned (Fig. 1c), from which we infer that the high-velocity outflow is moving at $> 50 \text{ km s}^{-1}$. This limit is consistent with the observed three-dimensional motions of Herbig–Haro objects in BN/KL^{14,23} and with the Doppler velocities of thermal CO and SiO emission¹¹. We note that the SiO maser emission from the limbs of the cones suggests material flowing away from the observer towards the northwest, opposite to the direction expected given inclination of the torus. However, we consider that the masers trace gas selectively. For small inclinations, optimal amplification can occur with the opposite sense of Doppler shift (that is, red- versus blue-shift) from gas along the axes of the cones, depending on the radial and angular dependence of the outflow velocity.

We conclude that the equatorial region is dominated by outflowing material and not by (rotating) natal material. We may extend the known radius of this outflow up to 1,000 AU by considering the previously mapped distribution of $v = 0$ SiO emission^{8,24}, to which three components contribute: high-velocity thermal, low-velocity thermal, and low-velocity maser emission. Both low-velocity components overlie the H₂O maser distribution in velocity and sky position within 350 AU of source I. The brightness distribution is asymmetric with respect to source I, as is the H₂O maser torus. On the smaller side of the torus, stronger SiO emission coincides with a richer distribution of H₂O masers, which is consistent with a density gradient in the surrounding medium, as seen at wavelengths of 8.7 and $12.4 \mu\text{m}$ (refs 3, 4). Although the apparent mixing of H₂O maser and SiO emission is difficult to understand (because silicate-grain formation depletes gas-phase SiO), we note that similar mixing is apparent in the envelopes of late-type stars^{25,26}.

There are three propitious alignments of structure that strengthen the case for the model proposed here. First, no H₂O masers lie within the high-velocity flow demarcated by the SiO masers (Fig. 1c). This flow is probably too hot or too turbulent to support maser action. Second, the four components of IRc2²⁷ shown in Fig. 2 subtend the opening angle of the northern outflow cone. They may have a direct line of sight to the protostar and may be heated by it or by interaction with the high-velocity flow. Infrared radiation from the southeasterly outflow is blocked by extinction in a foreground cloud core^{15,21}. Third, the magnetic field direction inferred from

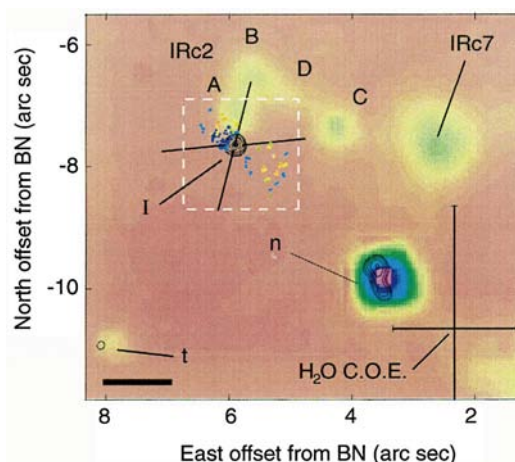


Figure 2 A near-infrared image (at $3.8 \mu\text{m}$ wavelength) of the region near IRc2²⁷, with contours representing 8.4-GHz emission⁵ and the H₂O maser distribution superposed. The bar indicates 1,000 AU. Labels mark the four components of IRc2 (A, B, C and D) and other infrared sources. Specifically, source n is the brightest infrared object in the field. It is probably a young massive star with an associated H II region¹⁵, though no dust emission have been detected yet at 3-mm wavelength¹⁵. The dashed box indicates the field of view in Fig. 1b; the 'X' largely within the box, highlights a possible relationship between the proposed outflow from source I and IRc2. Also shown is the centre of expansion (COE) for the low-velocity (18 km s^{-1}) flow in BN/KL, estimated in part from the proper motions of H₂O masers outside the field shown¹². A false-colour scale indicates the intensity of infrared emission, increasing from yellow to magenta.

dust emission at wavelengths of $100 \mu\text{m}$ and 3 mm is parallel to the high-velocity outflow axis^{28,29}. This suggests that the formation of the protostar and the alignment of the outflow axis may have been affected by the broad field distribution in OMC-1.

We argue that the masers clustered around source I provide evidence for a two-component outflow from a young star. In the context of star-formation theory for high-mass stars, no current model readily explains wide-angle collimation of a high-velocity bipolar outflow within tens of astronomical units of a protostar and a simultaneous low-velocity equatorial outflow. Magnetohydrodynamic mechanisms such as X-winds³⁰ and disk winds³¹ may be reasonable candidates. Alternatively, a spherical stellar wind that mixes with and ablates an (unseen) accretion disk up to ~ 350 AU in radius might generate a slow, dense equatorial wind, but this would not explain the observed conical structures. In addition, the presence of an H II region may preclude a disk. In this case, if accretion and outflow processes are linked, then the birth of an H II region could dissipate the observed equatorial outflow in about 100–200 years. On the basis of the alignment of the outflows from source I with other structures in BN/KL, we suggest that source I drives a substantial portion of the bulk motion in BN/KL; but we note that other sources, such as n (Fig. 2), rightly draw attention^{4,5} in this context. We anticipate that VLBA maps of SiO and H₂O maser proper motions and SiO maser polarizations could clarify the structure, dynamics and magnetic-field geometry in this region, which would help to resolve the debate about sources of energy and outflow in BN/KL. □

Received 13 October; accepted 19 November 1998.

- Kleinmann, D. E. & Low, F. J. Discovery of an infrared nebula in Orion. *Astrophys. J.* **149**, L1–L4 (1967).
- Genzel, R. & Stutzki, J. The Orion molecular cloud and star-forming region. *Annu. Rev. Astron. Astrophys.* **27**, 41–85 (1989).
- Gezari, D. Y. Mid-infrared imaging of Orion BN/KL: astrometry of IRc2 and the SiO maser. *Astrophys. J.* **396**, L43–L47 (1992).
- Gezari, D. Y. & Backman, D. E. 4.8–20 micron imaging of Orion BN/KL: II. A new look at luminosity sources and the role of IRc2. *Astrophys. Space Sci.* **224**, 45–52 (1995).
- Menten, K. M. & Reid, M. J. What is powering the Orion Kleinmann-Low infrared nebula? *Astrophys. J.* **445**, L157–L160 (1995).
- Churchwell, E., Wood, D. O. S., Felli, M. & Massi, M. Solar system-sized condensations in the Orion nebula. *Astrophys. J.* **321**, 516–519 (1987).

7. Plambeck, R. L., Wright, M. C. H. & Carlstrom, J. E. Velocity structure of the Orion-IRc2 SiO maser: evidence for an 80 AU diameter circumstellar disk. *Astrophys. J.* **348**, L65–L68 (1990).
8. Wright, M. C. H., Plambeck, R. L., Mundy, L. G. & Looney, L. W. SiO emission from a 1,000 AU disk in Orion KL. *Astrophys. J.* **455**, L185–L188 (1995).
9. Barvainis, R. The polarization of the SiO masers in Orion: maser emission from a rotating, expanding disk? *Astrophys. J.* **279**, 358–362 (1984).
10. Thronson, H. A. Jr *et al.* The Orion star-forming region: far-infrared and radio molecular observations. *Astron. J.* **91**, 1350–1356 (1986).
11. Wright, M. C. H., Plambeck, R. L. & Wilner, D. J. A multilane aperture synthesis study of Orion-KL. *Astrophys. J.* **469**, 216–237 (1996).
12. Genzel, R., Reid, M. J., Moran, J. M. & Downes, D. Proper motions and distances of H₂O maser sources. I. The outflow in Orion-KL. *Astrophys. J.* **244**, 884–902 (1981).
13. Allen, D. A. & Burton, M. G. Explosive ejection of matter associated with star formation in the Orion nebula. *Nature* **363**, 54–56 (1993).
14. Jones, B. F. & Walker, M. F. Proper motions of Herbig-Haro objects. VI. The M42 HH objects. *Astron. J.* **90**, 1320–1323 (1985).
15. Plambeck, R. L., Wright, M. C. H., Mundy, L. G. & Looney, L. W. Subarcsecond-resolution 86 GHz continuum maps of Orion KL. *Astrophys. J.* **455**, L189–L192 (1995).
16. Morita, K.-I., Hasegawa, T., Ukita, N., Okumura, S. K. & Ishiguro, M. Accurate positions of SiO masers in active star forming regions: Orion-KL, W51-IRS2, and Sgr-B2 MD5. *Publ. Astron. Soc. Jpn* **44**, 373–380 (1992).
17. Torrelles, J. M. *et al.* Radio continuum-H₂O maser systems in NGC 2071: H₂O masers tracing a jet (IRS 1) and a rotating proto-planetary disk of radius 20 AU (IRS 3). *Astrophys. J.* **505**, 756–765 (1998).
18. Zensus, J. A., Diamond, P. J. & Napier, P. J. (eds) *Very Long Baseline Interferometry and the VLBA* (Astron. Soc. Pacific, San Francisco, 1995).
19. Gaume, R. A., Wilson, T. L., Vrba, F. J., Johnston, K. J. & Schmid-Burgk, J. Water masers in Orion. *Astrophys. J.* **493**, 940–949 (1998).
20. Sugai, H. *et al.* Velocity field of the Orion-KL region in molecular hydrogen emission. *Astrophys. J.* **442**, 674–678 (1995).
21. Migena, V., Johnston, K. J., Pauls, T. A. & Wilson, T. L. The distribution and kinematics of ammonia in the Orion-KL nebula: high-sensitivity VLA maps of the NH₃(3,2) line. *Astrophys. J.* **347**, 294–301 (1989).
22. Greenhill, L. J. *et al.* in *The Impact of VLBI on Astrophysics and Geophysics* (eds Reid, M. J. & Moran, J. M.) 253–254 (Kluwer, Dordrecht, 1988).
23. Hu, X. Kinematic studies of Herbig-Haro objects in the Orion nebula. *Astron. J.* **112**, 2712–2717 (1996).
24. Chandler, C. J. & De Pree, C. G. Vibrational ground-state SiO $J = 1-0$ emission in Orion IRc2 imaged with the VLA. *Astrophys. J.* **455**, L67–L71 (1995).
25. Lucas, R. *et al.* Interferometric observations of SiO $v = 0$ thermal emission from evolved stars. *Astron. Astrophys.* **262**, 491–500 (1992).
26. Danchi, W. C., Bester, M., Degiacomi, C. G., Greenhill, L. J. & Townes, C. H. Characteristics of dust shells around 13 late-type stars. *Astron. J.* **107**, 1469–1513 (1994).
27. Dougados, C., Lena, P., Ridgway, S. T., Christou, J. C. & Probst, R. G. Near-infrared imaging of the Becklin-Neugebauer-IRc2 region in Orion with subarcsecond resolution. *Astrophys. J.* **406**, 112–121 (1993).
28. Schleuning, D. A. Far-infrared and submillimeter polarization of OMC-1: evidence for magnetically regulated star formation. *Astrophys. J.* **493**, 811–825 (1998).
29. Rao, R., Crutcher, R. M., Plambeck, R. L. & Wright, M. C. H. High-resolution millimeter-wave mapping of linearly polarized dust emission: magnetic field structure in Orion. *Astrophys. J.* **502**, L75–L78 (1998).
30. Shu, F. H., Najita, J., Ostriker, E. C. & Shang, H. Magnetocentrally driven flows from young stars and disks. V. Asymptotic collimation into jets. *Astrophys. J.* **455**, L155–L158 (1995).
31. Wardle, M. & Königl, A. The structure of protostellar accretion disks and the origin of bipolar flows. *Astrophys. J.* **410**, 218–238 (1993).
32. Garay, G., Moran, J. M. & Haschick, A. D. The Orion-KL super water maser. *Astrophys. J.* **338**, 244–261 (1989).

Acknowledgements. We thank M. Reid for assistance in recovering his VLA data, and C. Dougados and K. Menten for permission to use the infrared and radio images presented in Fig. 2. The National Radio Astronomy Observatory is a facility of the US NSF operated under cooperative agreement by Associated Universities, Inc. C.S. was supported by the NSF REU program. C.R.G. was supported by the NSF Stellar Astronomy and Astrophysics program.

Correspondence and requests for materials should be addressed to L.J.G. (e-mail: greenhill@cfa.harvard.edu).

Experimental observation of relativistic nonlinear Thomson scattering

Szu-yuan Chen, Anatoly Maksimchuk & Donald Umstadter

Center for Ultrafast Optical Science, University of Michigan, Ann Arbor, Michigan 48109, USA

Classical Thomson scattering¹—the scattering of low-intensity light by electrons—is a linear process, in that it does not change the frequency of the radiation; moreover, the magnetic-field component of light is not involved. But if the light intensity is extremely high ($\sim 10^{18}$ W cm⁻²), the electrons oscillate during the scattering process with velocities approaching the speed of light. In this relativistic regime, the effect of the magnetic and electric fields on the electron motion should become comparable, and the effective electron mass will increase. Consequently, electrons in such high fields are predicted to quiver nonlinearly, moving in

figure-of-eight patterns rather than in straight lines. Scattered photons should therefore be radiated at harmonics of the frequency of the incident light^{2–8}, with each harmonic having its own unique angular distribution^{4–6}. Ultrahigh-peak-power lasers⁹ offer a means of creating the huge photon densities required to study relativistic, or ‘nonlinear’ (ref. 6), Thomson scattering. Here we use such an approach to obtain direct experimental confirmation of the theoretical predictions of relativistic Thomson scattering. In the future, it may be possible to achieve coherent^{10,11} generation of the harmonics, a process that could be potentially utilized for ‘table-top’ X-ray sources.

We used a laser system that produces 400-fs-duration laser pulses at 1.053- μ m wavelength with a maximum peak power of 4×10^{12} W (4 TW). The laser beam, 50 mm in diameter, was focused with an $f/3.3$ parabolic mirror onto the front edge of a supersonic helium gas jet. The focal spot consisted of a gaussian spot with full-width at half-maximum (FWHM) of 7 μ m (containing 60% of the total energy) and a large ($>100 \mu$ m) dim spot. The helium gas was fully ionized by the foot of the laser pulse. A half-wave plate was used to rotate the axis of linear polarization of the laser beam in order to vary the azimuthal angle (ϕ) of observation. We define $\theta = 0^\circ$ as along the direction opposite to that of the laser propagation, and $\phi = 0^\circ$ as along the axis of linear polarization. In a linearly polarized laser field, electrons move in a figure-eight trajectory lying in the plane defined by the axis of linear polarization and the direction of beam propagation.

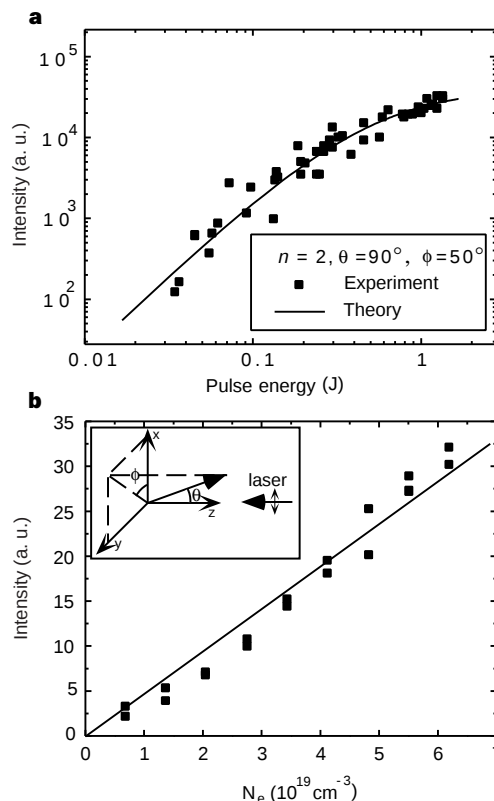


Figure 1 Dependence of the second-harmonic light on laser intensity and plasma density. Shown is the intensity of the second-harmonic light at $\theta = 90^\circ$, $\phi = 50^\circ$; **a**, as a function of laser pulse energy at 6.2×10^{19} cm⁻³ electron density, and **b**, as a function of plasma electron density (N_e) at 0.8-J laser pulse energy. (For 1-J laser pulse energy, the laser intensity is 4.4×10^{18} W cm⁻² and a_0 (see text) is 1.88.) Each data point represents the result of a single laser shot. The theoretical predictions for zero drift velocity are plotted as solid lines, for comparison. The only fitting parameter is a constant for normalization in all figures. The laser pulse energy refers to the energy in the main focal spot. The inset shows the coordinate system used.

Although the observation of harmonics in laser–plasma (or laser–electron-beam) interactions has been made by several groups^{12–16}, that alone is insufficient to identify unambiguously nonlinear Thomson scattering and its underlying dynamics. Several other mechanisms might generate continuum or harmonics under our experimental conditions, and, therefore, need to be isolated and discriminated from the signal generated by nonlinear Thomson scattering: (1) continuum generated from self-phase-modulation of the laser beam in the gas, (2) harmonics generated from atomic nonlinear susceptibility of the gas or, especially, from the ionization process¹⁷, (3) continuum generated either from (relativistic) self-phase-modulation of laser pulses in the plasma, or from electron–electron bremsstrahlung and electron–ion bremsstrahlung, and (4) harmonics generated from the interaction of laser pulses with a transverse electron-density gradient¹⁴.

The main focal spot of the laser pulse undergoes relativistic-ponderomotive self-channelling when high laser power and high gas density are used¹⁸. (The ponderomotive force is generated by the gradient of the optical pressure of the laser beam). Side-imaging ($\theta = 90^\circ$) of the first-harmonic light (at the laser frequency) from nonlinear Thomson scattering shows that the laser channel has a diameter of $<10\ \mu\text{m}$ FWHM. However, interferograms¹⁸ show that the diameter of the plasma column is about $100\text{--}200\ \mu\text{m}$, which is created by the wings with intensities $>10^{15}\ \text{W cm}^{-2}$ (the ionization threshold). Therefore, the light generated from laser–gas interactions should be observed to originate from the entire region of plasma, rather than from the narrow laser channel. Results of side-imaging ($\theta = 90^\circ$ and arbitrary ϕ) of the second and third harmonics using a matching interference filter (10-nm bandwidth) show that the signal is emitted only from the narrower laser channel.

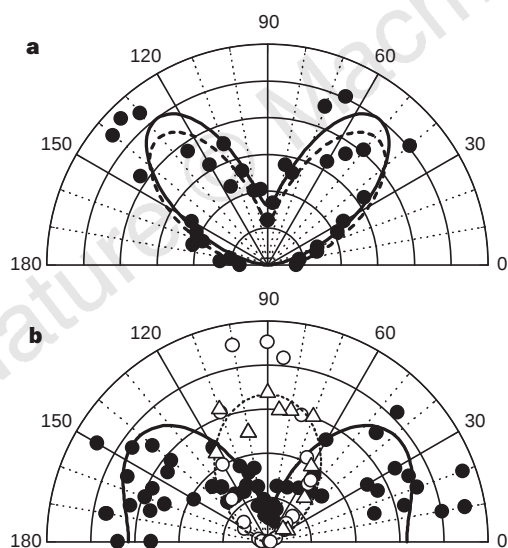


Figure 2 Angular pattern of the second-harmonic light. Shown are polar plots of the intensity of the second-harmonic light as a function of azimuthal angle (ϕ , in degrees) for 0.8-J pulse energy and $6.2 \times 10^{19}\ \text{cm}^{-3}$ electron density at $\theta = 90^\circ$ (**a**) and $\theta = 51^\circ$ (**b**). The intensity is in arbitrary units. Filled circles, experimental data; solid and dashed lines, theoretical results for zero and nonzero drift velocity ($v = 0.2c$) in the laser propagation direction, respectively. Open circles and triangles, experimental data for the first-harmonic signal taken at two different runs under the same conditions; dotted line, the theoretical result. Its dipole radiation pattern (peaked at $\phi = 90^\circ$) confirms that there is no depolarization effect in the plasma and that the collective effect of plasmas on the angular pattern is not significant (at least outside a narrow cone along the axis of laser propagation). Although the data for $\phi = 180^\circ$ to 360° are not plotted, they should be a mirror image of the data for $\phi = 0^\circ$ to 180° because of the intrinsic symmetry of the laser field, as predicted theoretically. Such symmetry has been verified in the experiment.

In addition, the images of the harmonics have spatial distributions similar to the images of the first-harmonic light, and their profiles vary in the same way as the laser power and gas density are changed. This rules out the possibility that the harmonic signal observed in the side images is a result of laser–gas interaction (mechanisms (1) and (2) above).

According to theory^{4–6}, the harmonic signal generated from nonlinear Thomson scattering should have two important features. First, it is linearly proportional to the electron density because it is an incoherent single-electron process (the harmonics generated from a collection of electrons interfere with each other destructively, leaving only an incoherent signal, which is equal to the single-electron results multiplied by the total number of electrons which radiate). Second, it increases roughly as I^n , where n is the harmonic number, and gradually saturates when a_0 is of the order of unity⁶. Here $a_0 = eE/m_0\omega_0c = 8.5 \times 10^{-10} \lambda I^{1/2}$ is the normalized vector potential, where E is the amplitude of laser electric field, $I = cE^2/8\pi$ (in W cm^{-2}) is the laser intensity, λ is the laser wavelength (in μm), e is the electron charge, m_0 is the electron mass, ω_0 is the laser frequency, and c is the speed of light. These are characteristically different from the behaviour of any other mechanisms. For instance, bremsstrahlung radiation should be proportional to the square of gas density ($N_e N_e$ or $N_e N_i$) where N_e is the electron density and N_i is the ion density. In the experiment reported here, the intensity of the harmonic signal was determined from the peak intensity or the average intensity of the images of harmonics, when it was plotted as a function of the observing angle, gas density and laser power. Both showed the same variations. Figure 1 shows the variation of the second-harmonic signal as a function of laser power and plasma (electron) density. The experimental results show a reasonable match to the theoretical predictions. The first and third harmonics show the same match with the theory.

Although the above two observations are consistent with nonlinear Thomson scattering as the source of the harmonic signal, observations of the unique angular patterns are necessary to prove that the detailed dynamics of nonlinear Thomson scattering are indeed the same as the theoretical prediction. Figure 2a shows the ϕ -dependence of the second-harmonic signal at $\theta = 90^\circ$. The experimental results match qualitatively the theoretical prediction, both having a quadrupole-type radiation pattern which is characteristically different from the dipole pattern for other mechanisms (1)–(4), and linear Thomson scattering. Other measurements, such as the ϕ -dependence of the second-harmonic light at $\theta = 51^\circ$ (an ‘anti-dipole’ pattern; Fig. 2b), and the ϕ -dependence of the third-harmonic light at $\theta = 90^\circ$ (a ‘butterfly’ pattern; Fig. 3), were also made, all showing reasonable matches between the experimental

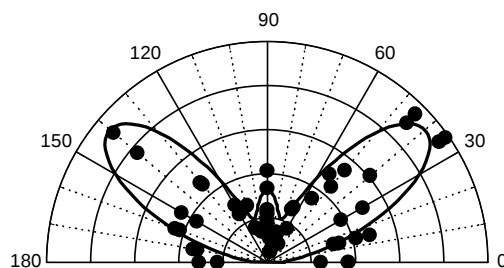


Figure 3 Angular pattern of the third-harmonic light. Shown is a polar plot of the intensity of the third-harmonic light as a function of azimuthal angle (ϕ , in degrees) at $\theta = 90^\circ$ for 0.8-J pulse energy and $6.2 \times 10^{19}\ \text{cm}^{-3}$ electron density. The intensity is in arbitrary units. Filled circles, experimental data; solid line, the theoretical result for zero drift velocity. The angular patterns of harmonics should not be sensitive to variation of laser intensity, as expected from theory and checked in the experiment. This is crucial to the success of our measurements because it alleviates the error caused by fluctuations of laser intensity.

data and the theoretical predictions. Such angular radiation patterns directly prove the electrons do indeed oscillate with a figure-of-eight trajectory in an intense (relativistic) laser field. The angular pattern of the first-harmonic light (linear component) of nonlinear Thomson scattering is also included in Fig. 2b for comparison.

The spectra of the second and the third harmonics each contain a peak at roughly the harmonic wavelength and a red-shifted broader peak, as shown in Fig. 4. The red-shifted broader peaks are believed to be part of the harmonic spectra generated by nonlinear Thomson scattering, because they vary in amplitude proportionally with the corresponding unshifted harmonic signals when the gas density and laser power are changed. It was expected that the spectra of the harmonics should be broadened tremendously for electrons in a high-fluid-velocity plasma wave⁵. A fast-phase-velocity electron plasma wave (with a maximum fluid velocity of $\sim 0.2c$) excited by stimulated Raman forward scattering¹⁹ was observed in this experiment at the highest laser power and gas density. But the spectral distribution of the harmonics was not observed to change significantly with variation of gas density and laser power, whereas the plasma wave amplitude did change with these variations. This indicates that such spectral structure has nothing to do with the collective drift motion of electrons in the plasma waves.

Although the angular radiation patterns of the harmonics could also be affected by such a $0.2c$ fluid-velocity oscillation, the changes are not significant enough (compare the solid and dashed lines in Fig. 2a) to be identified from the experimental data⁶. In other words, all measurements done in this experiment match qualitatively with the prediction of incoherent nonlinear Thomson scattering of electrons without drift motion; the results appear not to be affected by the existence of plasma waves, probably due to destructive coherent interference. The absolute scattering efficiency was measured to be 8×10^{-4} and 1×10^{-4} photons per electron per pulse for the second and third harmonics (including both the unshifted and red-shifted spectral components), respectively; these efficiencies were measured at $\theta = 90^\circ$ and $\phi = 50^\circ$, for an angle of collection of 7×10^{-3} sr. These numbers match reasonably well with the theoretical predictions for incoherent nonlinear Thomson scattering, which are respectively 8×10^{-4} and 5×10^{-4} photons per electron per pulse. □

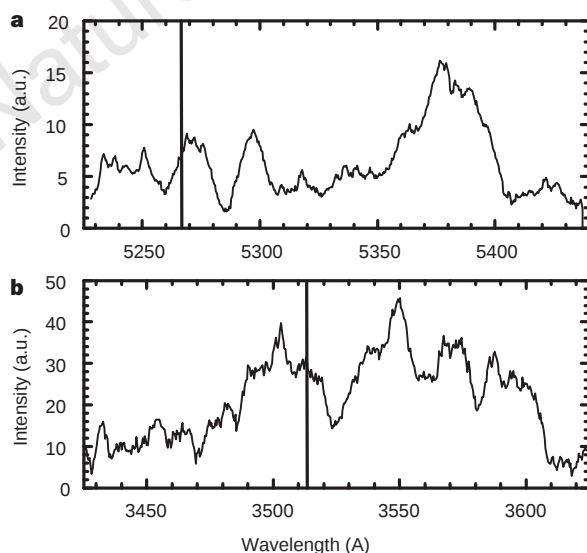


Figure 4 Spectra of harmonics. Shown are spectra of the second (a) and third (b) harmonics at $\theta = 90^\circ$, $\phi = 50^\circ$ for 0.8-J pulse energy and $6.2 \times 10^{19} \text{ cm}^{-3}$ electron density. Vertical lines indicate the wavelengths of the unshifted second and third harmonics. The intensities are plotted in arbitrary units. The spectra do not change with variation of ϕ at any specific θ , so the angular distributions measured are not affected by the choice of filter bandwidth.

Received 28 May; accepted 23 September 1998.

1. Thomson, J. J. *Conduction of Electricity through Gases* (Cambridge Univ. Press, 1906).
2. Vachaspati Harmonics in the scattering of light by free electrons. *Phys. Rev.* **128**, 664–666 (1962).
3. Brown, L. S. & Kibble, T. W. B. Interaction of intense laser beams with electrons. *Phys. Rev.* **133**, A705–A719 (1964).
4. Sarachik, E. S. & Schappert, G. T. Classical theory of the scattering of intense laser radiation by free electrons. *Phys. Rev. D* **1**, 2738–2753 (1970).
5. Castillo-Herrera, C. I. & Johnston, T. W. Incoherent harmonic emission from strong electromagnetic waves in plasmas. *IEEE Trans. Plasma Sci.* **21**, 125–135 (1993).
6. Esarey, E., Ride, S. K. & Sprangle, P. Nonlinear Thomson scattering of intense laser pulses from beams and plasmas. *Phys. Rev. E* **48**, 3003–3021 (1993).
7. Hartemann, F. V. & Luhmann, N. C. Jr Classical electrodynamical derivation of the radiation damping force. *Phys. Rev. Lett.* **74**, 1107–1110 (1995).
8. Hartemann, F. V. High-intensity scattering processes of relativistic electrons in vacuum. *Phys. Plasmas* **5**, 2037–2047 (1998).
9. Maine, P. et al. Generation of ultrahigh peak power pulses by chirped pulse amplification. *IEEE J. Quantum Electron.* **24**, 398–403 (1988).
10. Esarey, E. et al. Nonlinear analysis of relativistic harmonic generation by intense lasers in plasmas. *IEEE Trans. Plasma Sci.* **21**, 95–104 (1993).
11. Esarey, E. & Sprangle, P. Generation of stimulated backscattered harmonic generation from intense-laser interactions with beams and plasmas. *Phys. Rev. A* **45**, 5872–5882 (1992).
12. Meyer, J. & Zhu, Y. Second harmonic emission from an underdense laser-produced plasma and filamentation. *Phys. Fluids* **30**, 890–895 (1987).
13. Basov, N. G. et al. Investigation of $2\omega_0$ -harmonic generation in a laser plasma. *Sov. Phys. JETP* **49**, 1059–1067 (1979).
14. Malka, V. et al. Second harmonic generation and its interaction with relativistic plasma waves driven by forward Raman instability in underdense plasmas. *Phys. Plasmas* **4**, 1127–1131 (1997).
15. Englert, T. J. & Rinehart, E. A. Second-harmonic photons from the interaction of free electrons with intense laser radiation. *Phys. Rev. A* **28**, 1539–1545 (1983).
16. Bula, C. et al. Observation of nonlinear effects in Compton scattering. *Phys. Rev. Lett.* **76**, 3116–3119 (1996).
17. Brunel, F. Harmonic generation due to plasma effects in a gas undergoing multiphoton ionization in the high-intensity limit. *J. Opt. Soc. Am. B* **7**, 521–526 (1990).
18. Chen, S.-Y. et al. Evolution of a plasma waveguide created during relativistic-ponderomotive self-channeling of an intense laser pulse. *Phys. Rev. Lett.* **80**, 2610–2613 (1998).
19. Le Blanc, S. P. et al. Temporal characterization of a self-modulated laser wakefield. *Phys. Rev. Lett.* **77**, 5381–5384 (1996).

Acknowledgements. We thank G. Mourou, R. Wagner and X.-F. Wang for discussions. The work of S.C. and A.M. was supported by US NSF and that of D.U. by the Office of Energy Research, US DOE.

Correspondence and requests for materials should be addressed to D.U. (e-mail: dpu@umich.edu).

Electrostatically driven charge-ordering in Fe_2OBO_3

J. P. Attfield^{*†}, A. M. T. Bell^{*}, L. M. Rodriguez-Martinez^{*†}, J. M. Greneche[‡], R. J. Cernik[§], J. F. Clarke^{||} & D. A. Perkins^{||}

^{*}Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

[†]IRC in Superconductivity, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK

[‡]Laboratoire de Physique de l'Etat Condensé, UPRESA CNRS 6087, Université du Maine, 72085 Le Mans, France

[§]Synchrotron Radiation Source, CLRC Daresbury Laboratory, Warrington WA4 4AD, UK

^{||}Chemical Crystallography Laboratory, University of Oxford, Parks Road, Oxford, UK

Charge-ordering is an important phenomenon in conducting metal oxides: it leads to metal–insulator transitions¹ in manganite perovskites (which show ‘colossal’ magnetoresistances), and the Verwey² transition in magnetite (in which the material becomes insulating at low temperatures when the conduction electrons freeze into a regular array). Charge-ordered ‘stripes’ are found in some manganites^{3,4} and copper oxide superconductors⁵; in the latter case, dynamic fluctuations of the stripes have been proposed⁶ as a mechanism of high-temperature superconductivity. But an important unresolved issue is whether the charge-ordering in oxides is driven by electrostatic repulsions between the charges (Wigner crystallization⁷), or by the strains arising from electron–lattice interactions (such as Jahn–Teller distortions) involving different localized electronic states. Here we report measurements on iron oxoborate, Fe_2OBO_3 , that support the electrostatic repulsion charge-ordering mechanism: the system adopts a charge-ordered state below 317 K, in which Fe^{2+}

and Fe^{3+} ions are equally distributed over structurally distinct Fe sites. In contrast, the isostructural manganese oxoborate, Mn_2OBO_3 , has been previously shown⁸ to undergo charge-ordering through Jahn–Teller distortions. We therefore conclude that both mechanisms occur within the same structural arrangement.

The naturally occurring iron oxide magnetite, Fe_3O_4 , was one of the first mixed-valent materials found to possess unusual electronic properties^{2,9}. The charge distribution in this inverse spinel is written as $(\text{Fe}^{2+}\text{Fe}^{3+})\text{Fe}^{3+}\text{O}_4$ because the bracketed Fe^{2+} and Fe^{3+} are found within chains of edge-sharing FeO_6 octahedra interconnected by the additional, tetrahedrally coordinated Fe^{3+} . Octahedral site Fe^{2+} and Fe^{3+} spins order ferromagnetically below the Curie temperature $T_C = 858$ K, which facilitates spin-conserving transfer of the 'extra' electron from Fe^{2+} to Fe^{3+} , resulting in an appreciable conductivity of $\sim 10^2 \Omega^{-1} \text{cm}^{-1}$. This double-exchange mechanism is also responsible for the conducting ferromagnetic state of mixed-valent manganese oxide perovskites, which display large negative magnetoresistances¹ and the same effect has been observed in polycrystalline Fe_3O_4 owing to intergranular transport of the spin-polarized electrons¹⁰. At the Verwey transition in Fe_3O_4 ($T_V = 120$ K), a sharp decrease in conductivity is accompanied by a complex lattice distortion as a result of $\text{Fe}^{2+}/\text{Fe}^{3+}$ charge-ordering.

Selective replacement of the tetrahedral Fe^{3+} in Fe_3O_4 by boron gives the oxoborate $\text{Fe}^{2+}\text{Fe}^{3+}\text{OBO}_3$, which has a different, warwickite-type^{11,12} structure (Fig. 1) because boron bonds to only three of the four oxygen atoms. This phase was reported previously¹², but the electronic and magnetic properties have not been explored. We have obtained polycrystalline Fe_2OBO_3 from a sealed tube reaction of Fe metal, Fe_2O_3 and FeBO_3 at 1,100 °C and find evidence for charge-ordering from physical measurements.

Two phase transitions have been found in Fe_2OBO_3 . Low-temperature neutron diffraction¹³ (3–190 K on diffractometer D2B at ILL, Grenoble; see Supplementary information) and magnetization measurements (4–300 K in fields of 0.05, 0.5 and 5.0 T on a SQUID magnetometer; see Supplementary information) show that Fe_2OBO_3 orders ferrimagnetically below $T_C = 155$ K. The magnetic structure is almost antiferromagnetic (Fig. 1) but the magnetization measurements reveal a small ferromagnetic moment of $\sim 0.03 \mu_B$

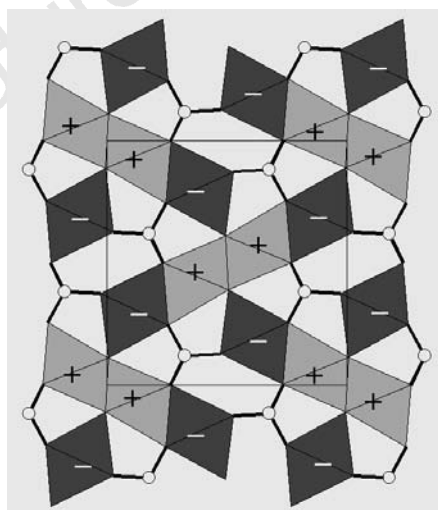


Figure 1 The Fe_2OBO_3 structure projected on the y - z plane (y vertical, z horizontal). Infinite chains of edge-sharing FeO_6 octahedra perpendicular to this plane form four octahedron wide 'ribbons' by sharing edges with neighbouring chains. An important feature is the presence of two structurally distinct chains of octahedra, $\text{Fe}(1)\text{O}_6$ and $\text{Fe}(2)\text{O}_6$, indicated by light and dark shading respectively. Below $T_C = 155$ K, the moments within each chain are ferromagnetically ordered in the directions indicated by plus or minus signs¹³. The saturated moment is $3.7 \mu_B$ per Fe atom.

per Fe atom in a 0.05 T field.

A second transition, assigned to charge-ordering (CO) for reasons outlined below, is found at $T_{CO} = 317$ K. A structural transition from monoclinic $P2_1/c$ symmetry to orthorhombic $Pm\bar{c}n$ symmetry occurs close to the mid-point of a broad semiconductor–semiconductor transition, at which the conductivity drops by a factor of ~ 3 (Fig. 2). The conductivity is proportional to $e^{-E_a/kT}$ above and below the transition, although the limited data ranges do not exclude other functions, and the estimated activation energies, E_a , are 0.31 eV above 350 K and 0.35 eV below 270 K. The changes in conductivity and structure are small (the unit-cell parameters vary from $a = 3.16879(8) \text{ \AA}$, $b = 9.3835(3) \text{ \AA}$, $c = 9.2503(3) \text{ \AA}$, $\beta = 90.220(1)^\circ$ at 3 K, to $a = 3.1779(1) \text{ \AA}$, $b = 9.3945(1) \text{ \AA}$, $c = 9.2495(1) \text{ \AA}$, $\beta = 90^\circ$ at 337 K), but the dramatic change in the ^{57}Fe Mössbauer spectra at around 320 K (Fig. 3) shows that charge localization occurs at the transition.

A striking feature of the low-temperature monoclinic Fe_2OBO_3 structure is an equal distribution of Fe^{2+} and Fe^{3+} over the two structurally distinct octahedral Fe sites (that is, the charge distribution is $\text{Fe}(1)_{0.5}^{2+}\text{Fe}(1)_{0.5}^{3+}\text{Fe}(2)_{0.5}^{2+}\text{Fe}(2)_{0.5}^{3+}\text{OBO}_3$). Although the two types of FeO_6 octahedra (Fig. 1) are both very distorted, with Fe–O bond lengths varying between 1.92 and 2.23 Å (from the 3 K neutron refinement), the mean $\text{Fe}(1)\text{--O}$ and $\text{Fe}(2)\text{--O}$ distances of 2.085(2) and 2.082(2) Å are equal within their errors. Comparing these to the expected $\text{Fe}^{2+}\text{--O}$ and $\text{Fe}^{3+}\text{--O}$ distances of 2.16 and 2.02 Å (ref. 14) enables the deviation of the charge at each site from the average 2.5+ value to be estimated as ≤ 0.01 . This is supported by the very small difference between the saturated magnetic

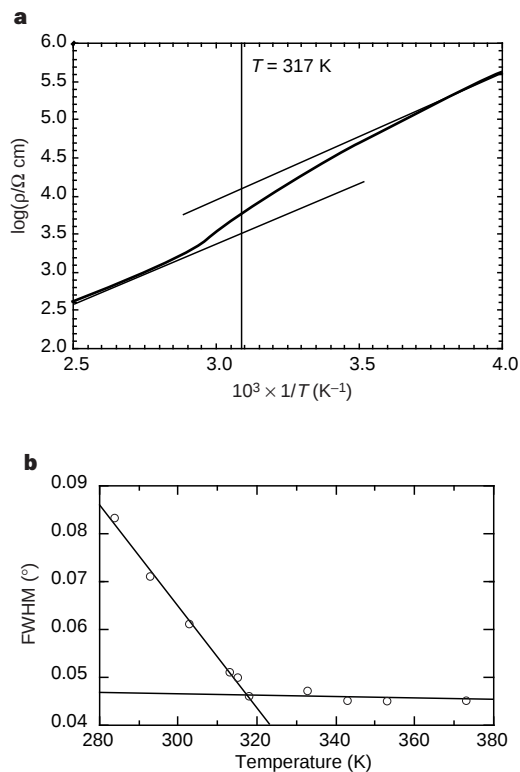


Figure 2 Conductivity and structural anomalies at the charge-ordering transition in Fe_2OBO_3 . **a**, Log(resistivity) plotted against inverse temperature, with linear high and low temperature limits and the temperature of the structural transition at 317 K shown. Resistivity was measured by a four-probe method from 412 to 240 K, below which the sample resistance was too large to be measured accurately. The structural transition was found from the thermal variation of the full width at half maximum (FWHM) of the $102/\bar{1}02$ doublet (**b**), which collapses to a single peak at the transition, measured by synchrotron powder X-ray diffraction on station 9.1 at the Synchrotron Radiation Source, Daresbury, UK (see Supplementary Information).

moments at the two sites, which is attributable to their inequivalent structural environments. Furthermore, the Mössbauer spectra between 160 and 270 K consist of four doublets of equal intensity, corresponding to equal amounts of Fe^{2+} and Fe^{3+} , each localized at the Fe(1) and Fe(2) sites.

This exact equipartition of divalent and trivalent ions over the two cation sites is not found in the many known mixed-metal warwickites such as FeMnOBO_3 (ref. 15) and MgScOBO_3 (ref. 16). In addition, the mixed-valent Mn analogue has a charge distribution $\text{Mn}(1)^{3+}\text{Mn}(2)^{2+}\text{OBO}_3$, as evidenced by a Jahn–Teller distortion at the Mn(1) site, and mean Mn(1)–O and Mn(2)–O distances of 2.064 and 2.179 Å, respectively⁸. Nor is charge equipartition found in the structurally related iron oxoborate, $\text{Fe}_2^{2+}\text{Fe}^{3+}\text{O}_2\text{BO}_3$, which shows an unequal charge distribution over four octahedral Fe sites¹⁷. Therefore, the distribution in Fe_2OBO_3 results both from strong electrostatic repulsions between the extra Fe^{2+} electrons (that surpass the electron–lattice interactions) and from there being half as many electrons as available sites, which leads to a regular charge-ordered state corresponding to a Wigner crystal of extra electrons on the Fe sites. Mn_2OBO_3 represents the opposite extreme, in which predominant Jahn–Teller electron–lattice interactions due to Mn^{3+} lead to a complete segregation of the charges between the two octahedral sites according to their ease of distortion. We conclude that the charge-ordered states of the mixed-valent iron oxides Fe_2OBO_3 and Fe_3O_4 are Wigner crystals, whereas those in Mn^{3+} oxides such as Mn_2OBO_3 and manganite perovskites are ‘distortion ordered’. This difference arises from the strong Jahn–Teller interaction for Mn^{3+} which does not occur for Fe^{2+} or Fe^{3+} .

The Fe ions are on planes of mirror symmetry in the high-temperature orthorhombic Fe_2OBO_3 structure (Fig. 4a). Any long-range charge-ordering such as the simple alternating scheme shown

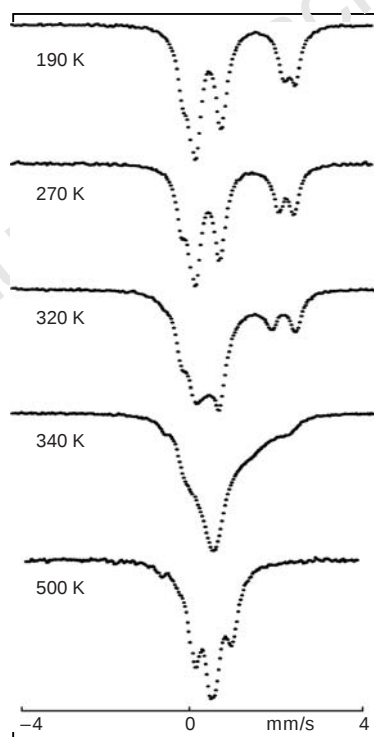


Figure 3 Mössbauer spectra of Fe_2OBO_3 between 190 and 500 K, collected in transmission geometry on a conventional constant acceleration spectrometer. Below 270 K, these consist of two Fe^{2+} and two Fe^{3+} doublets of equal intensity, but between 270 and 380 K the signals broaden and coalesce owing to the exchange of Fe^{2+} and Fe^{3+} environments through electron hopping. Well defined Mössbauer peaks above 380 K are consistent with rapid electron hopping at two structurally distinct Fe^{2+} sites.

in the lower part of Fig. 4b destroys the mirror symmetry and the repulsions lead to a tilting of the ribbons of octahedra, consistent with the observed enlargement of the β angle below T_{CO} . However, the a -axis periodicity should also increase by a factor of two (or some larger integer in a more complex charge-ordering scheme), which is not observed in our X-ray or neutron diffraction experiments. This suggests that the extent of the charge-ordered domains is small (~ 10 – 100 Å), so that any superstructure peaks are too weak and broad to be observed against the background in our diffraction patterns. This may reflect a large concentration of defects, such as the anti-phase boundary shown in Fig. 4b, that preserve the direction of the monoclinic distortion, but do not propagate the coherent doubling of the lattice periodicity. Hence, the observed long-range monoclinic lattice distortion can arise from charge-order within much smaller domains, which have been termed ‘Wigner nanocrystals’¹⁸. We note that distinct Fe^{2+} and Fe^{3+} sites are not observed either in the low-temperature structure of Fe_3O_4 (ref. 19), so the long-range structural distortion may again result from nanoscale charge-ordering. Wigner crystallization thus appears to be too easily disrupted to give long-range structural order, whereas ‘distortion-ordered’ materials such as $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$ can show well defined charge-ordered superstructures.

Our results show that low-symmetry mixed-valent materials with complex structures, such as the warwickite-type oxoborates, provide insight into the properties of high-symmetry oxides such as the spinel magnetite. Both Fe_2OBO_3 and Fe_3O_4 undergo a magnetic transition at which ferromagnetic ordering occurs within octahedral ($\text{Fe}^{2+}\text{Fe}^{3+}$) chains, and a structural transition associated with a decrease in electronic conductivity. The differences between the transition temperatures of the two materials stem from their different structures and the replacement of paramagnetic Fe^{3+} by diamagnetic B^{3+} . In Fe_3O_4 , strong superexchange interactions between the octahedral ($\text{Fe}^{2+}\text{Fe}^{3+}$) and tetrahedral Fe^{3+} spins result in a high magnetic ordering temperature ($T_{\text{C}} = 858$ K) and ferromagnetic alignment of the octahedral Fe^{2+} and Fe^{3+} moments. This leads to a facile transfer from Fe^{2+} to Fe^{3+} of the minority spin electrons, enabling the conducting state to persist to a relatively low temperature until charge-order occurs at the Verwey transition, $T_{\text{V}} = 120$ K. Weaker, frustrated, superexchange pathways in Fe_2OBO_3 lead to a much lower magnetic ordering temperature ($T_{\text{C}} = 155$ K), so that electron hopping in the paramagnetic region is less facile and charge-ordering occurs at a relatively high

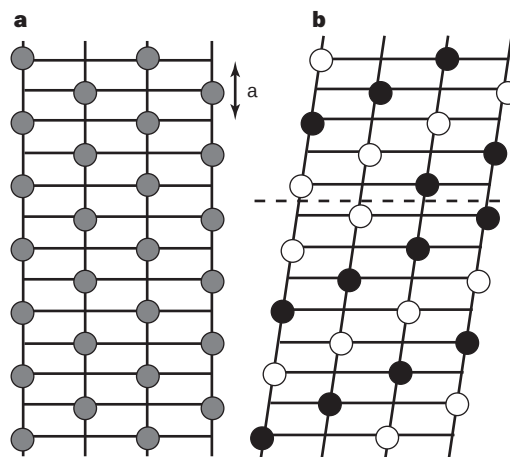


Figure 4 Charge disorder and order in Fe_2OBO_3 . **a**, The arrangement of sites within a ribbon of four Fe site chains in orthorhombic Fe_2OBO_3 above T_{CO} . **b**, A possible ordering scheme of Fe^{2+} and Fe^{3+} ions below T_{CO} also showing an antiphase boundary that preserves the sense of the monoclinic distortion.

$T_{\text{CO}} = 317$ K. Hence a conducting, magnetically ordered state is not found in Fe_2OBO_3 as $T_{\text{C}} < T_{\text{CO}}$. The differences between the estimated activation energies for electron hopping in the charge-disordered ($E_{\text{a}} \approx 0$ for $T > T_{\text{V}}$) and ordered ($E_{\text{a}} \approx 0.04$ eV for $T < T_{\text{V}}$)¹³ states of magnetite and those of Fe_2OBO_3 ($E_{\text{a}} \approx 0.31$ eV for $E_{\text{a}} \approx 0.35$ eV for $T > T_{\text{CO}}$) are equal, showing that this small difference is essentially independent of spin alignment, although the ferromagnetic order in magnetite reduces both activation energies by 0.3 eV relative to those in paramagnetic Fe_2OBO_3 . □

Received 7 May; accepted 20 August 1998.

1. Rao, C. N. R. & Cheetham, A. K. Giant magnetoresistance, charge-ordering, and related aspects of manganates and other oxide systems. *Adv. Mater.* **9**, 1009–1017 (1997).
2. Verwey, E. J. W. Electronic conduction of magnetite (Fe_3O_4) and its transition point at low temperatures. *Nature* **144**, 327–328 (1939).
3. Radaelli, P. G., Cox, D. E., Marezio, M. & Cheong, S. W. Charge, orbital, and magnetic ordering in $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$. *Phys. Rev. B* **55**, 3015–3023 (1997).
4. Mori, S., Chen, C. H. & Cheong, S. W. Pairing of charge-ordered stripes in $(\text{La}, \text{Ca})\text{MnO}_3$. *Nature* **392**, 473–476 (1998).
5. Tranquada, J. M., Sternlieb, B. J., Axe, J. D., Nakamura, Y. & Uchida, S. Evidence for stripe correlations of spins and holes in copper oxide superconductors. *Nature* **375**, 561–563 (1995).
6. Salkola, M. L., Emery, V. J. & Kivelson, S. A. Implications of charge-ordering for single-particle properties of high- T_{c} superconductors. *Phys. Rev. Lett.* **77**, 155–158 (1996).
7. Mott, N. F. In *Metal–Insulator Transitions* 2nd edn 213–215 (Taylor & Francis, London, 1990).
8. Norrestam, R., Kritikos, M. & Sjoerdin, A. Manganese(II,III) oxyborate, Mn_2OBO_3 : a distorted homometallic warwickite—synthesis, crystal structure, band calculations, and magnetic susceptibility. *J. Solid State Chem.* **114**, 311–316 (1995).
9. Honig, J. M. Analysis of the Verwey transition in magnetite. *J. All. Comp.* **229**, 24–39 (1995).
10. Coey, J. M. D., Berkowitz, A. E., Balcells, L., Putris, F. F. & Parker, F. T. Magnetoresistance of magnetite. *Appl. Phys. Lett.* **72**, 734–736 (1998).
11. Takeuchi, Y., Watanabe, T. & Ito, T. The crystal structures of warwickite, ludwigite and pinakolite. *Acta Crystallogr.* **3**, 98–107 (1950).
12. Bertaut, E. F. Structures des boroferrites. *Acta Crystallogr.* **3**, 473–474 (1950).
13. Attfield, J. P., Clarke, J. F. & Perkins, D. A. Magnetic and crystal structures of iron borates. *Physica B* **180**, 581–584 (1992).
14. Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr. A* **32**, 751–767 (1976).
15. Blumh, K. & Utzolano, A. Synthesis and crystal structure of manganese borate oxides— $\text{MnFe}(\text{BO}_3)_2\text{O}$ and $\text{MnAl}_{0.5}\text{Y}_{0.5}(\text{BO}_3)_2\text{O}$. *Z. Naturforsch. B* **50**, 1146–1150 (1995).
16. Norrestam, R. Structural investigation of 2 synthetic warwickites—undistorted orthorhombic $\text{Mg}_2\text{ScOBO}_3$ and distorted monoclinic $\text{Mg}_{0.76}\text{Mn}_{1.24}\text{OBO}_3$. *Z. Krist.* **189**, 1–11 (1989).
17. Swinnea, J. S. & Steinfink, H. Crystal structure and Mössbauer spectrum of vonsenite, $2\text{FeO} \cdot \text{FeBO}_3$. *Am. Mineral.* **5**, 827–832 (1983).
18. Rozenberg, G. K., Hearn, G. R., Paternak, M. P., Metcalf, P. A. & Honig, J. M. Nature of the Verwey transition in magnetite (Fe_3O_4) to pressures of 16 GPa. *Phys. Rev. B* **53**, 6482–6487 (1996).
19. Izumi, M. et al. Structure of magnetite (Fe_3O_4) below the Verwey transition temperature. *Acta Crystallogr. B* **38**, 2121–2133 (1982).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank EPSRC for provision of beam time at SRS and ILL, and G. Bushnell-Wye (SRS) and A. Hewat (ILL) for assistance with data collection. Support for A.M.T.B. from EPSRC and the Daresbury Laboratory and for L.M.R.M. from the Basque Government is acknowledged.

Correspondence and requests for materials should be addressed to J.P.A. (e-mail: jpa14@cam.ac.uk).

Spin-triplet superconductivity in Sr_2RuO_4 identified by ^{17}O Knight shift

K. Ishida*, H. Mukuda*, Y. Kitaoka*, K. Asayama*, Z. Q. Mao†‡, Y. Mori† & Y. Maeno†‡

* Department of Physical Science, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan

† Department of Physics, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan

‡ CREST, Japan Science and Technology Corporation, Kawaguchi, Saitama 332-0012, Japan

Superconductivity—one of the best understood many-body problems in physics—has again become a challenge following the discovery of unconventional superconducting materials: these include heavy-fermion¹, organic² and the high-transition-temperature copper oxide³ superconductors. In conventional superconductors, the electrons form superconducting Cooper

pairs in a spin-singlet state, which has zero total spin ($S = 0$). In principle, Cooper pairs can also form in a spin-triplet state ($S = 1$), analogous to the spin-triplet 'p-wave' state of paired neutral fermions in superfluid ^3He (ref. 4). At present, the heavy-fermion compound UPt_3 is the only known spin-triplet superconductor^{5,6}, although the layered oxide superconductor Sr_2RuO_4 (ref. 7) is believed, on theoretical grounds⁸, to be a promising candidate. The most direct means of identifying the spin state of Cooper pairs is from measurements of their spin susceptibility, which can be determined by the Knight shift (as probed by nuclear magnetic resonance (NMR)). Here we report Knight-shift measurements of Sr_2RuO_4 using ^{17}O NMR. Our results show no change in spin susceptibility on passing through the superconducting transition temperature, which provides the definitive identification of Sr_2RuO_4 as a spin-triplet superconductor.

Nuclear magnetic resonance and nuclear quadrupole resonance (NMR/NQR) probe superconducting and normal-state properties through measurements of nuclear spin–lattice relaxation rate $1/T_1$ and Knight shift which measures the effective field at the nucleus produced by the electrons. The BCS theory predicted that $1/T_1$ exhibits a peak just below the superconducting transition temperature T_{c} (ref. 9). This coherence peak is characteristic of, and essential for, the spin-singlet s-wave pairing realized in ordinary metals in which a weak-coupling model is valid. However, $1/T_1$ of Ru in Sr_2RuO_4 measured at zero field did not show the coherence peak (ref. 10), giving a strong argument against the s-wave picture. This result is reinforced by the observation of the extremely strong suppression of T_{c} by disorder introduced by non-magnetic impurities (ref. 11). The unconventional nature of the superconductivity in Sr_2RuO_4 is confirmed also by the recent observation of the broken time-reversal symmetry (ref. 12), an exotic property expected only for non-s-wave pairing.

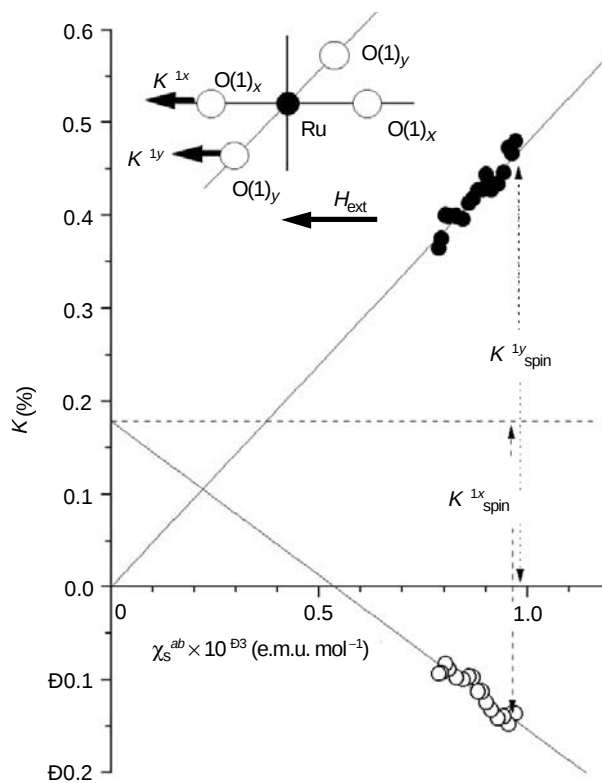


Figure 1 K - χ plot of Sr_2RuO_4 . The Knight shifts for the two in-plane oxygen sites, K^{1x} and K^{1y} , are plotted against the spin part of the bulk susceptibility, $\chi_s^{ab} (= \chi^{ab} - \chi_{\text{dia}} - \chi_{\text{orb}})$. Inset shows the definition of K^{1x} and K^{1y} .

For the spin-singlet pairing, the ratio of the spin susceptibility of the Cooper pair χ_s to that of the normal state χ_n is given by¹³;

$$\frac{\chi_s}{\chi_n} = -\frac{2}{N_0} \int_0^\infty N_s(E) \frac{\partial f(E)}{\partial E} dE \quad (1)$$

where N_0 and $N_s(E)$ are the quasiparticle density of states (DOS) in the normal state and in the superconducting state, respectively, and $f(E)$ is the Fermi distribution function. For the s -wave pairing, the DOS is given by $N_s(E) = N_0|E|/\sqrt{E^2 - \Delta^2}$ (for $|E| > \Delta$) and $N_s = 0$ (for $|E| < \Delta$), in which $\Delta(T/T_c)$ is the superconducting energy gap, and equation (1) reduces to the Yosida function $Y(T/T_c)$. At low T , χ_s decreases exponentially as $\exp(-\Delta/k_B T)$. In a spin-singlet d -wave state, the decay of χ_s is not exponential. As $N_s(E)$ is proportional to the energy E for $|E| \ll \Delta$ for the $d_{x^2-y^2}$ symmetry as in high- T_c copper oxides, χ_s is proportional to T at low T (refs 14, 15). As long as the pairing is spin-singlet, χ_s diminishes to zero as $T \rightarrow 0$.

For a spin-triplet p -wave superconductor in the tetragonal, quasi-two-dimensional lattice like Sr_2RuO_4 , the most stable states in the weak-coupling approximation are the equal-spin pairing states⁸. If one can neglect the spin-orbit coupling, χ_s in these states remains constant in any field direction even below T_c . On the other hand, if the spins are pinned along an easy axis or plane owing to the crystal anisotropy, χ_s will be at least partially reduced below T_c when a magnetic field (H_{ext}) is applied perpendicular to that direction. This is the case for the spin-triplet pairing in UPT_3 in a hexagonal lattice⁶.

The electronic properties of Sr_2RuO_4 are dominated by the three quasi-two-dimensional bands, derived from hybridization of Ru $4d$ orbitals with O $2p$ orbitals (antibonding $d-p\pi$ orbitals)¹⁶. The effect of strong Coulomb repulsion among the electrons, reflected in the Fermi-liquid properties¹⁶, tends to favour pairing states with non-

zero orbital angular momentum. In fact, the spin-triplet pairing has been suggested rather than the spin-singlet pairing^{8,18}.

The Knight shift measurement below T_c by Ru NMR is technically difficult; owing to the very small gyromagnetic ratio of the Ru nucleus, $^{101}\gamma = 0.2193 \text{ MHz kOe}^{-1}$, the field allowed within the upper critical fields of $H_{c2lc}(0) = 0.75 \text{ kOe}$ and $H_{c2lab}(0) = 15 \text{ kOe}$ (see Fig. S1 in Supplementary information) is too small for the precise Ru NMR measurements. The Knight shift measurement in the superconducting state must therefore be carried out by ^{17}O NMR using the samples for which ^{17}O (with a nuclear spin $I = 5/2$) is substituted for ^{16}O (without a nuclear spin).

We report here data on two high-quality single crystals in the flat-plate shape, about $3 \times 6 \times 1 \text{ mm}$, grown by a floating-zone method. The a.c. susceptibility measurements indicated very sharp superconducting transitions with the onset at $T_c = 1.486 \text{ K}$ and 1.430 K . The transition width was $\Delta T_c = 52 \text{ mK}$ for both crystals. The substitution of ^{17}O for ^{16}O was carried out by annealing the crystals in $80\% \text{ } ^{17}\text{O}_2/20\% \text{ } ^{16}\text{O}_2$ gas at $1,050^\circ\text{C}$ for 14 days. The ratio of the NMR intensities indicates that the in-plane O(1) in the RuO_2 layer and the apical O(2) are nearly equally substituted.

We have successfully measured ^{17}O Knight shift for all the O sites. Knight shift in the temperature range $1.4\text{--}200 \text{ K}$ was measured at 58.1 MHz ($H_{\text{ext}} \approx 100 \text{ kOe}$); Knight shift in the $15 \text{ mK--}1.8 \text{ K}$ range was measured at 3.8 MHz ($H_{\text{ext}} \approx 6.5 \text{ kOe}$). We will focus below on the Knight shift of O(1) sites in a field parallel to the RuO_2 plane (the a - b plane) because of the relevance to the superconducting symmetry. When the external field is along the a axis (Ru-O-Ru bonding direction), O(1) sites are no longer equivalent; as shown in Fig. 1 inset, half of O(1) has the Ru-O bonding direction parallel to H_{ext} (O(1)_x) and the other half perpendicular to H_{ext} (O(1)_y). The difference in the electric-quadrupole interactions results in a clear

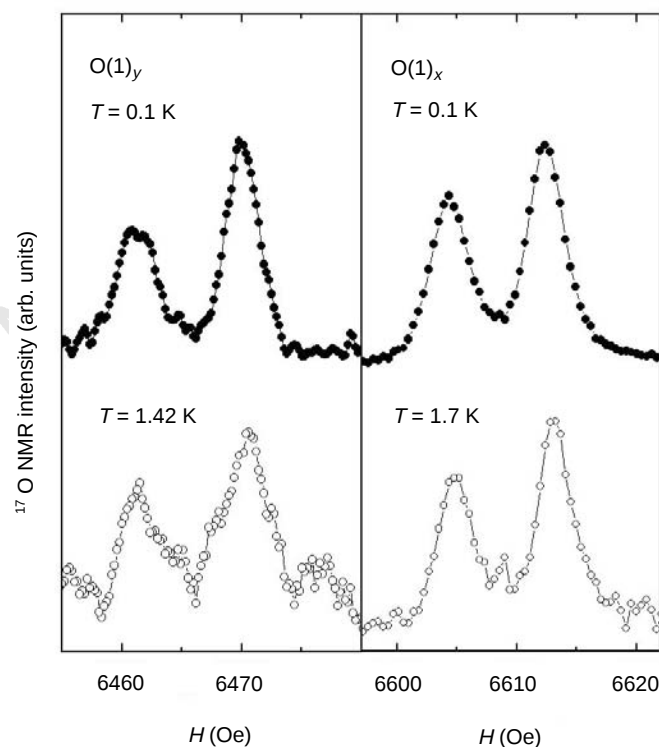


Figure 2 ^{17}O NMR spectra of O(1)_x and O(1)_y in both the normal (bottom) and the superconducting (top) state of Sr_2RuO_4 . A fast-Fourier-transform technique, at a frequency of 3.8 MHz , was used. Two peaks in each spectrum originate from a slight difference in the orientations of the two single crystals. Each peak corresponds to the central peak of the quintet. (We note that the spin parts of the Knight shift, $K_s^x \approx -0.3\%$ and $K_s^y \approx -0.5\%$, correspond to a shift of $+20 \text{ Oe}$ and -33 Oe , respectively.)

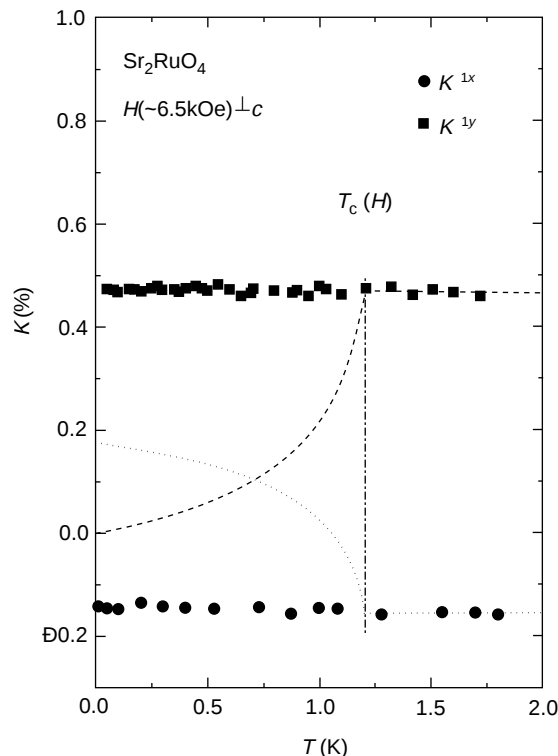


Figure 3 Temperature dependence of K^{1x} and K^{1y} at low temperatures. Broken lines below T_c indicate the calculation for the spin-singlet d -wave state in two dimensions with $d_{x^2-y^2}$ symmetry, using the parameters $\Delta(\phi) = \Delta_0 \cos(2\phi)$ and $2\Delta_0 = 8k_B T_c$ which are compatible with those of $\text{YBa}_2\text{Cu}_3\text{O}_7$ (ref. 15). ϕ is the angle on the cylindrical Fermi surface of the CuO_2 plane.

splitting of the NMR peaks of O(1) into two quintets. The Knight shift (K) at constant frequency (ν_0) is defined as the field difference between H_0 ($= \nu_0/\gamma_n$, where γ_n is the gyromagnetic ratio) and H_{res} (the resonance field), divided by H_{res} : $K = (H_0 - H_{\text{res}})/H_{\text{res}}$.

It is important to extract the spin part of the Knight shift from the plot of K^{1x} and K^{1y} against the spin part of the bulk susceptibility parallel to the a - b plane (χ_s^{ab}) with temperature as an implicit parameter (Fig. 1). The spin susceptibility in the superconducting state cannot be extracted from the bulk susceptibility as it is dominated by diamagnetism due to the Meissner effect. In contrast, the Knight shift, which originates from the hyperfine coupling to the polarized conduction electrons, remains, providing local information on their spin and orbital states. The bulk susceptibility χ^{ab} measured by a SQUID (superconducting quantum interference device) magnetometer in a field of 10 kOe is composed of the spin part $\chi_s^{ab}(T)$, the orbital part χ_{orb} , in which the contribution from Ru 4d orbit is $\chi_{\text{orb}}^{\text{Ad}} = 1.54 \times 10^{-4}$ (e.m.u. mol $^{-1}$) (ref. 10), as well as the diamagnetic susceptibility $\chi_{\text{dia}} = -0.91 \times 10^{-4}$ (e.m.u. mol $^{-1}$) (ref. 17) of the core electrons. Owing to the strong hybridization between Ru 4d(t_{2g}) and O- p_π orbitals, the spin contribution $K_s(T)$ in K^{1x} and K^{1y} is proportional to $\chi_s^{ab}(T)$. The observed K^i ($i = 1x$ and $1y$) is expressed as $K^i = K_s^i + K_{\text{orb}}^i = A^i \chi_s^{ab} + K_{\text{orb}}^i$, where A^i is the hyperfine coupling constants of the spin contribution, and K_{orb} originates from the orbital susceptibility from the O 2p orbitals. The large anisotropy in K^{1x} and K^{1y} indicates that the Knight shifts at O(1) sites are dominated by the dipolar field from the spin density on O 2p orbitals. The K - χ plot of Fig. 1 gives $K_{\text{orb}}^{1x} \approx 0.18\%$ and $K_{\text{orb}}^{1y} \approx 0.0\%$; and $K_s^{1x}(T) \approx -0.3\%$ and $K_s^{1y}(T) \approx 0.5\%$ at 1.5 K, which should be proportional to the spin susceptibility of quasiparticles to form the Cooper pairs at T_c .

Figure 2 shows the ^{17}O NMR spectra for the O(1) $_x$ and O(1) $_y$ sites, obtained by the Fourier-transform technique of spin-echo signals, both below and above $T_c(H_{\text{ext}}) \approx 1.2$ K for $H_{\text{ext}} = 6.5$ kOe parallel to the a axis. (See also Supplementary information, Fig. S2). We had stacked the two crystals with their a - b planes parallel to each other to check for any sample dependence of the superconducting properties. The separation of the peaks is due to the slight misalignment of the a -axis direction between the two crystals. By analysing the NMR spectra, however, we deduce that the deviation of H_{ext} from the a -axis direction within the a - b plane is less than 5° . The full-width at half-maximum (FWHM) is less than 5 Oe, guaranteeing a highly precise determination of the Knight shift. To check that the crystal became superconducting under the external field of $H_{\text{ext}} = 6.5$ kOe, we measured the temperature dependence of $1/T_1$ under the same condition (see Supplementary information, Fig. S3); $1/T_1/T$ increased gradually below ~ 1.3 K and exhibited a broad peak at 0.4 K. In contrast, for $H_{\text{ext}} = 6.5$ kOe parallel to the c -axis, which does destroy the superconductivity, $1/T_1/T$ remained constant to low temperatures. The observed peak appearing much below T_c is quite different from the behaviour of the coherence peak, and thus most probably originates from motions of vortices in the superconducting mixed state. In fact, essentially the same behaviour has been observed in organic superconducting compounds with quasi-two-dimensional SC characteristics¹⁹. These observations guarantee that the Knight shift measurements probed the spin susceptibility in the superconducting state.

Figure 3 shows the temperature dependence of K^{1x} and K^{1y} both above and below T_c for the crystal showing a slightly higher H_{res} in Fig. 2; the Knight shift for the other crystal was essentially the same. It is remarkable that both K^{1x} and K^{1y} are unchanged across T_c within the experimental precision. As the sample is in the clean limit, with a mean free path $l \approx 5,000$ Å (which is approximately an order of magnitude larger than the coherence length $\xi_{ab} \approx 660$ Å), any model based on impurity scattering effect fails to account for the apparent invariance of the spin susceptibility of Cooper pairs in the superconducting state²⁰. For comparison, the expected behaviour for a spin-singlet d -wave pairing is represented by dotted

curves in Fig. 3. For any spin-singlet pairing states with weak spin-orbit scattering, it is expected that the Knight shift approaches $K_{\text{orb}}^{1x} \approx 0.18\%$ and $K_{\text{orb}}^{1y} \approx 0\%$ as $T \rightarrow 0$. The experimental results clearly demonstrate no change of the spin susceptibility of Cooper pairs, and therefore lead to the conclusion that spin-triplet superconductivity is realized in Sr_2RuO_4 .

The pairing wavefunction of the spin-triplet state is most conveniently described in terms of the 'd-vector' (ref. 8). Among the possible spin-triplet states appropriate to Sr_2RuO_4 , one of the most promising, consistent with the present Knight shift results, is a p -wave state (with the orbital momentum $L = 1$) expressed by $\mathbf{d} = \hat{z}(\hat{k}_x \pm i\hat{k}_y)$, which breaks the time-reversal symmetry⁸. For this equal-pairing state, the Cooper pairs consist of the spin pairs $|\uparrow\uparrow\rangle$ and $|\downarrow\downarrow\rangle$ with their quantization axis perpendicular to the c -axis direction. As a result, an invariant χ_s is expected when the field is parallel to any direction within the a - b plane, in accordance with the observation. Although the diminishing χ_s/χ_n following the Yosida function $Y(T/T_c)$ is expected for the field parallel to the c -axis, the Knight shift measurement required to confirm this behaviour is technically not possible because $H_{c2||c}(0)$ is only 0.75 kOe.

We now compare the two spin-triplet superconductors. The Cooper-pair wavefunction of the heavy-fermion compound UPt_3 is much more complicated than a simple p -wave^{5,6}. Furthermore, the superconducting phase in UPt_3 is closely associated with a state of static antiferromagnetic ordering rather than a ferromagnetic ground state²¹. In contrast, the superconductivity in the layered oxide Sr_2RuO_4 is most probably of a simple p -wave type, and is closely associated with quasi-two-dimensional ferromagnetism. The two systems are very different in many respects; the work reported here should therefore promote further examination of the unusual characteristics of spin-triplet superconductivity from a more fundamental perspective. □

Received 8 September; accepted 21 October 1998.

- MacLaughlin, D. E. *et al.* Nuclear magnetic resonance and heavy-fermion superconductivity in (U,Th)Be $_{13}$. *Phys. Rev. Lett.* **53**, 1833–1836 (1984).
- Mayaffre, H. *et al.* Superconducting state of κ -(ET) $_2\text{Cu}[\text{N}(\text{CN})_2]\text{Br}$ studied by ^{13}C NMR: evidence for vortex-core-induced nuclear relaxation and unconventional pairing. *Phys. Rev. Lett.* **75**, 4122–4125 (1995).
- Kitaoka, Y. *et al.* Nuclear relaxation and Knight shift studies of copper in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. *J. Phys. Soc. Jpn* **57**, 30–33 (1988).
- Lee, D. M. The extraordinary phase of liquid ^3He . *Rev. Mod. Phys.* **69**, 645–665 (1997).
- Tou, H. *et al.* Odd-parity superconductivity with parallel spin pairing in UPt_3 : evidence from ^{195}Pt Knight shift study. *Phys. Rev. Lett.* **77**, 1374–1377 (1996).
- Tou, H. *et al.* Nonunitary spin-triplet superconductivity in UPt_3 : evidence from ^{195}Pt Knight shift study. *Phys. Rev. Lett.* **80**, 3129–3132 (1996).
- Maeno, Y. *et al.* Superconductivity in a layered perovskite without copper: Sr_2RuO_4 . *Nature* **372**, 532–534 (1994).
- Rice, T. M. & Sigrist, M. Sr_2RuO_4 : an electronic analogue of ^3He ? *J. Phys. Condens. Matter* **7**, L643–L648 (1995).
- Hebel, L. C. & Slichter, C. P. Nuclear spin relaxation in normal and superconducting aluminum. *Phys. Rev.* **113**, 1504–1519 (1959).
- Ishida, K. *et al.* Anisotropic pairing in superconducting Sr_2RuO_4 : Ru NMR and NQR studies. *Phys. Rev. B* **56**, R505–R508 (1997).
- Mackenzie, A. P. *et al.* Extremely strong dependence of superconductivity on disorder in Sr_2RuO_4 . *Phys. Rev. Lett.* **80**, 161–164 (1998).
- Luke, G. M. *et al.* Time-reversal symmetry breaking superconductivity in Sr_2RuO_4 . *Nature* **394**, 558–561 (1998).
- Yosida, K. Paramagnetic susceptibility in superconductors. *Phys. Rev.* **110**, 769 (1958).
- Barrett, S. E. *et al.* ^{63}Cu Knight shift in the superconducting state of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. *Phys. Rev. B* **41**, 6283–6296 (1990).
- Ishida, K. *et al.* Cu NMR and NQR studies of impurities-doped $\text{YBa}_2(\text{Cu}_{1-x}\text{M}_x)\text{O}_7$ ($\text{M} = \text{Zn}$ and Ni). *J. Phys. Soc. Jpn* **62**, 2803–2818 (1993).
- Oguchi, T. Electronic band structure of the superconductor Sr_2RuO_4 . *Phys. Rev. B* **51**, 1385–1388 (1995).
- Maeno, Y. *et al.* Two dimensional Fermi liquid behavior of the superconductor Sr_2RuO_4 . *J. Phys. Soc. Jpn* **66**, 1405–1408 (1997).
- Mazin, I. I. & Singh, D. J. Ferromagnetic spin fluctuation induced superconductivity in Sr_2RuO_4 . *Phys. Rev. Lett.* **79**, 733–736 (1997).
- Kanoda, K. NMR relaxation rate in the superconducting state of the organic conductor, $\kappa(\text{BEDT-TTF})_2\text{Cu}[\text{N}(\text{CN})_2]\text{Br}$. *Phys. Rev. B* **54**, 76–79 (1996).
- Anderson, P. W. Knight shift in superconductors. *Phys. Rev. Lett.* **3**, 325–356 (1959).
- de Visser, A. *et al.* UPt_3 , heavy fermions and superconductivity. *Physica B* **147**, 81–160 (1987).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank K. Miyake, O. Nariaki, M. Sigrist, K. Yamada, G.-q. Zheng and H. Tou, for discussions and comments, and T. Ando for magnetization measurements. This work was supported by the Grants-in-Aid for COE Research from the Ministry of Education, Sport, Science and Culture of Japan. One of us (H.M.) was supported by JPS research fellowships for young scientists.

Correspondence and requests for materials should be addressed to K.I. (e-mail: ishida@mp.es.osaka-u.ac.jp).

Polysaccharide elasticity governed by chair–boat transitions of the glucopyranose ring

Piotr E. Marszalek*, Andres F. Oberhauser*, Yuan-Ping Pang† & Julio M. Fernandez*

* Department of Physiology and Biophysics, † Mayo Clinic Cancer Center and Department of Pharmacology, Mayo Foundation, Rochester, Minnesota 55905, USA

Many common, biologically important polysaccharides contain pyranose rings made of five carbon atoms and one oxygen atom. They occur in a variety of cellular structures, where they are often subjected to considerable tensile stress^{1–6}. The polysaccharides are thought to respond to this stress by elastic deformation, but the underlying molecular rearrangements allowing such a response remain poorly understood. It is typically assumed, however, that the pyranose ring structure is inelastic and locked into a chair-like conformation. Here we describe single-molecule force measurements^{7–12} on individual polysaccharides that identify the pyranose rings as the structural unit controlling the molecule's elasticity. In particular, we find that the enthalpic component of the polymer elasticity^{10,11,13,14} of amylose, dextran and pullulan is eliminated once their pyranose rings are cleaved. We interpret these observations as indicating that the elasticity of the three polysaccharides results from a force-induced elongation of the ring structure and a final transition from a chair-like to a boat-like conformation. We expect that the force-induced deformation of pyranose rings reported here plays an important role in accommodating mechanical stresses and modulating ligand binding in biological systems.

Polymeric molecules based on the pyranose ring form a wide variety of polysaccharides that receive different names depending on the type of the monomer(s) and the glycosidic linkages. X-ray crystallography and neutron diffraction have shown that the most stable conformation of the pyranose ring structures is the chair ⁴C₁ conformer^{15–17}. Molecular dynamics simulations have confirmed that the chair ⁴C₁ conformer is the most populated structure: these simulations also suggested that the ring structure in the chair conformation is flexible because a range of chair distortions are allowed without significant changes in the conformational energy^{18–20}.

Some of these polysaccharides can be oxidized by periodate under mild conditions. The reaction is highly specific and results in the cleavage of the ring structure to form an acyclic backbone with intact glycosidic linkages²¹. We have found that such ring cleavage removes the enthalpic component of the elasticity of the three polysaccharides we have studied, demonstrating the central role played by the pyranose ring in polysaccharide elasticity. Figure 1 shows the effect, over time, of mild periodate oxidation on the elasticity of single dextran (Fig. 1a), amylose (Fig. 1b) and pullulan (Fig. 1c) molecules.

The dextran molecule is a polymer formed by glycosidic bonds linking the carbon atoms number 1 and 6 of consecutive α -D-glucopyranose rings (Fig. 1a inset). As previously shown by Rief *et al.*¹⁰, dextran molecules have a large enthalpic component in their elasticity that is evident as a prolonged plateau in the force–extension curve (Fig. 1a, black trace). The dextran monomer has two sites that are cleaved by periodate oxidation (2–3 and 3–4; Fig. 1a, red arrows in inset). As the figure shows, periodate oxidation causes a gradual disappearance of the enthalpic elasticity. The

elasticity of polymers such as dextran has been shown to be well described by a modified 'freely jointed chain' (FJC) model that predicts the relationship between the extension of the polymer and the entropic restoring force generated^{10,13,22}. The adjustable parameters of the FJC model are the contour length of the polymer, l_c , the Kuhn (segment) length, l_k , and the elasticity of each individual segment, S_e . Fits of the FJC function to the force–extension curves obtained for the polysaccharides adequately described their elasticity (thin lines, Fig. 1). The changes in the elasticity of the dextran molecule during periodate oxidation can be followed by using the FJC model. In the low-force regime, the native dextran molecule is well described by the FJC model with a Kuhn length of $l_k = 0.44$ nm (corresponding to the size of a monomer; see Table 1), and segment elasticity $S_e = 14,600 \pm 2,700$ pN nm^{–1} ($n = 23$). However, after periodate oxidation the Kuhn length becomes smaller, $l_k = 0.2 \pm 0.026$ nm ($n = 13$), and the segments become less extensible ($S_e = 34,200 \pm 8,300$ pN nm^{–1}, $n = 13$).

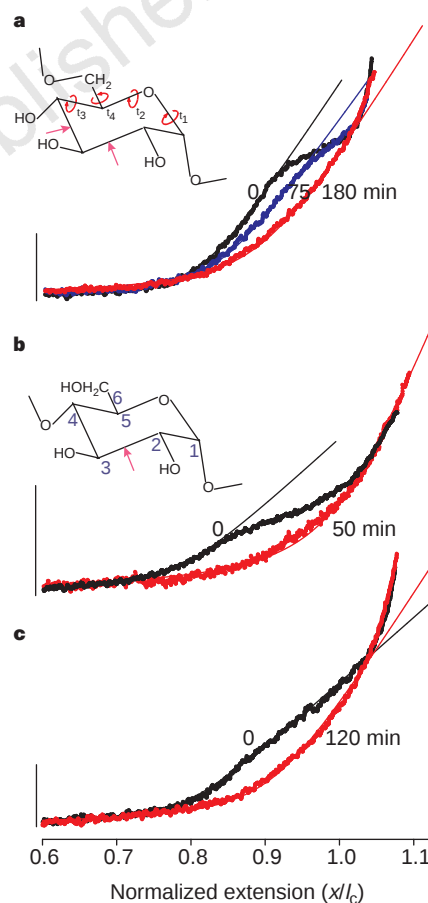


Figure 1 Force–extension curves for single polysaccharide molecules. These curves show that cleavage of the pyranose ring by mild periodate oxidation nearly eliminates the enthalpic component of the elasticity of these polymers. The insets in **a** and **b** show the chemical structure of the monomers of dextran and amylose, respectively. The red arrows mark the cleavage sites. t_1 – t_4 represent the ring torsion angles. Oxidation with 5 mM periodate for up to 180 min eliminated the plateau region of the force–extension curves of dextran (**a**), amylose (**b**) and pullulan (**c**) and altered their elasticity such that their force–extension curves approached those of a chain made of rigid elements. In all cases, the extension, x , was normalized by the contour length for each molecule, l_c . To monitor the changes in elasticity during oxidation, we fitted the experimental force–extension curves with the FJC model (thin lines). The FJC fits show that periodate oxidation reduced the Kuhn segment length and increased the stiffness of the segments (see text for details). These results demonstrate that the pyranose ring is the principal extensible element of these polysaccharides.

A large enthalpic elasticity is also observed in amylose, an α -(1 $\rightarrow$ 4)-D-glucopyranose polymer (Fig. 1b). However, the plateau observed in the force–extension curve for amylose has a much lower force threshold (275 ± 45 pN; $n = 11$) than that observed for dextran (850 ± 140 pN; $n = 23$). The amylose monomer has only one site that is cleaved by periodate oxidation, breaking the structure of the pyranose ring (2–3; Fig. 1b, red arrow in inset). The polysaccharide pullulan is another polymer of glucose combining the α -(1 $\rightarrow$ 4) and α -(1 $\rightarrow$ 6) glycosidic linkages (2:1 ratio) which also shows a pronounced enthalpic elasticity that appears to result from a combination of those observed for dextran and amylose (Fig. 1c). The pullulan monomers contain the combined cleavage sites of both dextran and amylose (not shown).

Similar to the results with dextran, periodate oxidation eliminated the enthalpic component of the chain elasticity of amylose and pullulan (Fig. 1b, c) and made their segments more rigid. Cleavage of the amylose ring reduced the length of the Kuhn segment from 0.45 nm (Table 1) to 0.18 ± 0.025 nm ($n = 12$) and increased segment elasticity (from $S_e = 5,600 \pm 800$ pN nm⁻¹ to $34,000 \pm 7,300$ pN nm⁻¹; $n = 12$). Cleavage of the rings in pullulan reduced the length of the Kuhn segment from 0.45 nm to 0.2 ± 0.015 nm ($n = 6$) and increased their segment elasticity (from $S_e = 10,200 \pm 930$ pN nm⁻¹ to $47,800 \pm 5,000$ pN nm⁻¹; $n = 6$). In all cases, periodate oxidation alters the chain segments such that they become shorter and stiffer.

The force–extension curves of dextran, amylose and pullulan are all significantly different, and seem to depend on the attachment points of the glycosidic bond to the glucopyranose ring (Fig. 1). However, cleavage of the pyranose ring converts the elastic behaviour of these polysaccharides into a simple entropic spring which is similar for all three molecules (Fig. 2).

We note that the force–extension curves reported here for amylose and pullulan are fully reversible and show a complete lack of hysteresis, similar to what was observed earlier for dextran¹⁰. This indicates that the enthalpic elasticity arises from the conformation of the monomers rather than from breaking secondary structure, as was proposed for the elasticity of DNA and xanthan, molecules that show large hysteresis during extension–contraction cycles^{13,23}.

A prominent characteristic of the enthalpic elasticity of amylose and dextran is a plateau region in the force–extension curve that indicates that the molecule elongates at approximately constant force. Our data show that this elongation must originate from the ring structure. Indeed, a six-membered ring structure can acquire several conformations: chair, boat, twist and half-twist²⁴. The glucopyranose ring is typically in the chair (⁴C₁) conformation. The free energy of interconversion between conformations ranges

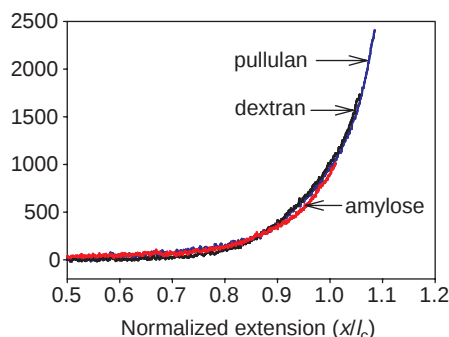


Figure 2 On cleavage of the pyranose ring, single molecules of amylose, pullulan and dextran display identical elastic behaviour. Ring cleavage converts these different polysaccharide chains into similar structures where all the bonds of the polymer backbone, including those that were components of the closed ring structure, can rotate and align under force, contributing equally to the (mainly) entropic elasticity of the chain.

up to 12 kcal mol⁻¹ for the free monomers in solution²⁵. The distance between two consecutive glycosidic oxygens (a residue vector, O_iO_j)—which determines the length of the monomer in a polysaccharide chain—is expected to vary with the pyranose ring conformation. (This is evident on inspection of a simple ball-and-stick model of the pyranose; Fig. 3a.) Stretching the polysaccharide chain with an external force is likely to favour conformations of the ring that provide a greater separation of the glycosidic oxygens, as this will lead to an extension of the molecule. In order to verify this view, we have made *ab initio* calculations of the length of the residue vector O_iO_j in all thermodynamically accessible conformations—including the chair and boat conformations of the pyranose ring for amylose and dextran (Table 1). For these molecules, the transition from chair to boat conformation increases the length of the O_iO_j initial calculations of the extension of the O_iO_j vector during the chair–boat transition for amylose (16.2%) and dextran (22.5%) closely match the experimental observations (Table 1; Fig. 3b).

Cellulose is a glucopyranose polymer identical to amylose except

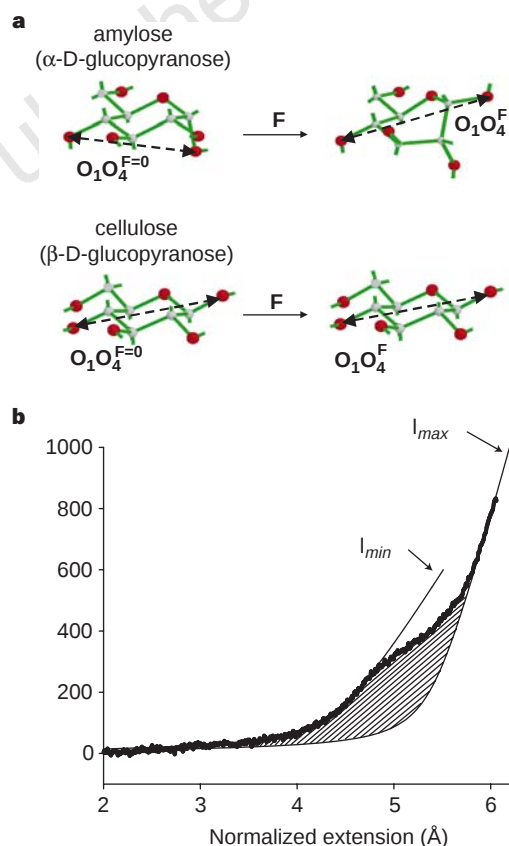


Figure 3 Increased separation of glycosidic oxygen atoms during chair–boat transitions of the glucopyranose ring explains the extensibility of amylose. **a**, The diagrams show that the distance between two consecutive glycosidic oxygens O_1 – O_4 increases during a force-driven chair–boat transition when the glycosidic oxygen O_1 is in the axial position (amylose). By contrast, the O_1 – O_4 distance is already at a maximum when O_1 is in the equatorial position (cellulose). *Ab initio* calculations of the separation of glycosidic oxygen atoms during chair–boat transitions (Table 1) accurately predict the extensibility of amylose, dextran and cellulose. **b**, A force–extension curve for amylose (thick trace) was fitted with the FJC model using a Kuhn length of 4.533 Å before the plateau and a Kuhn length of 5.408 Å after the plateau region (see Table 1). The computed fractional increase in contour length for this recording was 18%. The figure shows the data after normalization by $(O_1-O_4)/l_c$. The hatched area corresponds to the change in ΔG per monomer to convert it from the chair to the boat conformation. The free energy, ΔG , was determined to be 4.6 kcal mol⁻¹ (range 2.5–4.8 kcal mol⁻¹, $n = 11$). Similar measurements made with single dextran molecules showed a ΔG of ~ 11 kcal mol⁻¹.

that it has an equatorial glycosidic bond to the pyranose ring (β -(1 $\rightarrow$ 4) linkages). *Ab initio* calculations of the O_1O_4 vector for cellulose show that the length of this vector is already at a maximum in the chair conformation (Fig. 3a; Table 1). A transition to the boat and other conformations would decrease the O_1O_4 distance and is therefore prohibited in the molecule under a stretching force. Hence, these calculations predict that the cellulose molecule should behave as an entropic spring without showing the plateau region which characterizes the conformational transition of the ring structure. Indeed, recordings of the force-extension relationship for carboxymethylated cellulose (made using an atomic force microscope, AFM) do not show a measurable enthalpic component²³. Similarly, the force-extension relationship for methylcellulose, another soluble derivative of cellulose, does not have a measurable enthalpic component (not shown). These observations are consistent with the known rigidity of cellulose.

We have demonstrated two components of enthalpic elasticity in polysaccharides: under an applied force we observed a continuous elastic deformation of the pyranose ring (for example, 0–200 pN for amylose) and a transitional component (>200 pN). This latter component occurs when the pyranose ring adopts extended conformations that are normally less populated due to their higher conformational energies (chair-boat transition; Fig. 3a). The range of forces required to trigger these conformational changes are found in cellular systems, such as the bacterial and plant cell walls that must accommodate the high tensile stress generated by changes in turgor pressure^{5,6}. For example, we have found that pectin, a pyranose-based polysaccharide (an α -(1 $\rightarrow$ 4) D-galactouronan) and one of the main components of the plant cell wall, has a large enthalpic elasticity, similar to that of amylose. Single pectin molecules undergo an abrupt extension by 18% ($n = 7$) of their length at forces bigger than 200 pN (not shown). Hence, the structural assembly of different types of polysaccharides such as pectins and cellulose may allow plant cells to finely control the extensibility of their cell wall.

Hydroxyl and other functional groups attached to the pyranose ring coordinate the binding of lectins²⁶. It is interesting to consider that the force-induced ring deformations demonstrated here may alter the reactivity of the pyranose ring, allowing or preventing the binding of ligands.

Much of the recent progress in understanding the dynamics of biological molecules, a central focus of modern biology, comes from the manipulation of single molecules. The direct mechanical manipulation of a pyranose ring, as demonstrated here, gives a unique and unusual way to control ring conformations which are conventionally governed by the nature of the substituents attached to the rings. As demonstrated by Barton²⁷, conformational analysis of ring-based substrate and ligands such as steroids provided the elements needed to understand their biological action²⁷. Here, we expand the boundaries of conformational analysis by including force-driven changes in ring structure as important, but hitherto unrecognized, biological events. □

Table 1 *Ab initio* calculation of the separation of glycosidic oxygen atoms during chair-boat transitions

	Calculated				Experimental $\frac{l_{\max} - l_{\min}}{l_{\max}}$ (%)
	$O_i-O_j^{F=0}$ (Å)	$O_i-O_j^F$ (Å)	Δ (Å)	$\frac{\Delta}{O_i-O_j^F}$ (%)	
Amylose $_{\alpha-1,4}$	4.533	5.408	0.875	16.2	16.8 ± 1.8 ($n = 11$)
Dextran $_{\alpha-1,6}$	4.412	5.696	1.284	22.5	19.5 ± 2.1 ($n = 23$)
Cellulose $_{\beta-1,4}$	5.417	5.417	0	0	0 ($n = 29$)

This table compares the distance between glycosidic oxygen atoms, in the relaxed ($F = 0$) and stretched conformations, $(O_i-O_j^{F=0} - O_i-O_j^F)/O_i-O_j^F$, with the experimentally measured extension of the contour length, $(l_{\max} - l_{\min})/l_{\max}$. The experimental values for $l_{\min}$ and $l_{\max}$ were obtained from the fits of the FJC model to the force-extension curve before and after the plateau region (thin lines in Fig. 3b).

Methods

Single-molecule atomic force microscopy. Our custom-made AFM apparatus, as well as its mode of operation, have been described elsewhere¹²: it is capable of measuring individual molecules. The spring constant of each individual AFM tip was calibrated in solution, using the equipartition theorem as described by Florin *et al.*²⁸. This method gives values for the spring constant of the cantilever which are within 20% of the values obtained by other methods²⁸. The polysaccharides used—dextran T500 (Pharmacia Biotech); amylose (type III: from potato; Sigma); pullulan (from *Aureobasidium pullulans*; Sigma) and methylcellulose (Sigma)—are linear or >95% linear (dextran) homopolymers of D-glucopyranose (relative molecular mass $M_r > 10^5$). Dextran has short branches which frequently consist of just a single glucose at the position 3. Pectin (from citrus fruits; Sigma) is an α -(1 $\rightarrow$ 4) D-galactouronan. Dextran, pullulan, methylcellulose and pectin were dissolved in water at a concentration of 0.1–10% (w/v). Amylose was solubilized by wetting with ethanol (100 mg ml⁻¹), followed by treatment with sodium hydroxide solution (10%) and heating. The final concentration of amylose was 0.2–1% (w/v). A layer of polysaccharide molecules was created by drying a drop of these solutions onto glass coverslips followed by extensive rinsing. This procedure leaves a monolayer of polysaccharide molecules tightly adsorbed to the glass surface²³. The measurements were carried out in water or in PBS buffer (pectin). For ring cleavage, the polysaccharide samples were treated with 5 mM sodium meta-periodate (Sigma) for up to 180 min. To pick polysaccharide molecules, an AFM tip was pressed down onto the sample for 1–3 s at forces of several nanonewtons. A subsequent force-extension measurement typically revealed a complex pattern with many overlapping force peaks. By manipulating the polysaccharide concentration and the force and duration of the tip-sample contact, it was possible to regularly obtain single polysaccharide force-extension curves.

***Ab initio* calculation of the glucopyranose ring conformations.** Different conformations of the glucopyranose ring were generated by specifying all discrete possibilities at increments of 60 degrees of arc in the range 0–360° for torsions t_1 – t_4 (Fig. 1a). These conformers were optimized first with molecular mechanics calculation (Quanta97/CHARMm)²⁹ and then with *ab initio* calculation (GAUSSIAN94, HF/6-31G*/HG/6-31G*)³⁰. The length of the residue vector in the relaxed state of a polysaccharide (no external force) was obtained from the distance between the two consecutive glycosidic oxygen atoms (O_i – O_j) in the most stable glucopyranose conformer (chair). The corresponding length of the residue vector in the stretched polysaccharide was obtained from the maximal O_i – O_j distance found in all the conformers optimized by the *ab initio* calculation. The separations of the glycosidic oxygen atoms O_i – O_j in the conformations representing the relaxed ($F = 0$) and stretched states of the glucopyranose monomers in amylose, dextran, and cellulose are shown in Table 1. For amylose and dextran, these maximal distances corresponded to the boat conformation. For cellulose, the O_i – O_j distance is already maximal in the chair conformation indicating that an external force will not trigger a chair-boat transition.

Received 18 September; accepted 25 November 1998.

- McEver, R. P., Moore, K. L. & Cummings, R. D. Leukocyte trafficking mediated by selective carbohydrate interactions. *J. Biol. Chem.* **270**, 11025–11028 (1995).
- Rosen, S. D. & Bertozzi, C. R. Leukocyte adhesion: two selectins converge on sulphate. *Curr. Biol.* **6**, 261–264 (1996).
- Alon, R., Hammer, D. A. & Springer, T. A. Lifetime of the P-selectin-carbohydrate bond and its response to tensile force in hydrodynamic flow. *Nature* **374**, 539–542 (1995).
- Puri, K. D., Chen, S. & Springer, T. A. Modifying the mechanical property and shear threshold of I-selectin adhesion independently of equilibrium properties. *Nature* **392**, 930–933 (1998).
- Cosgrove, D. J. Relaxation in a high-stress environment: the molecular bases of extensible cell walls and cell enlargement. *Plant Cell* **9**, 1031–1041 (1997).
- Thwaites, J. J. & Mendelson, N. H. Biomechanics of bacterial walls: studies of bacterial thread made from *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **82**, 2163–2167 (1985).
- Bustamante, C., Rivetti, C. & Keller, D. J. Scanning force microscopy under aqueous solutions. *Curr. Opin. Struct. Biol.* **7**, 709–716 (1997).
- Florin, E. L., Moy, V. T. & Gaub, H. E. Adhesion forces between individual ligand-receptor pairs. *Science* **264**, 415–417 (1994).
- Lee, G. U., Chrissy, L. A. & Colton, R. J. Direct measurement of the forces between complementary strands of DNA. *Science* **266**, 771–773 (1994).
- Rief, M., Oesterhelt, F., Heymann, B. & Gaub, H. E. Single molecule force spectroscopy on polysaccharides by atomic force microscopy. *Science* **275**, 1295–1297 (1997).
- Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109–1112 (1997).
- Oberhauser, A. F., Marszalek, P. E., Erickson, H. P. & Fernandez, J. M. The molecular elasticity of the extracellular matrix protein tenascin. *Nature* **393**, 181–185 (1998).

13. Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271**, 795–799 (1996).
14. Rief, M., Fernandez, J. M. & Gaub, H. E. Elastically coupled two-level-systems as a model for biopolymer extensibility. *Phys. Rev. Lett.* **81**, 4764–4767 (1998).
15. Brown, G. M. & Levy, H. A. α -D-Glucose: precise determination of crystal and molecular structure by neutron-diffraction analysis. *Science* **147**, 1038–1039 (1965).
16. Chu, S. C. C. & Jeffrey, G. A. The refinement of the crystal structures of β -D-glucose and cellobiose. *Acta Crystallogr. B* **24**, 830–838 (1968).
17. Brown, G. M. & Levy, H. A. α -D-Glucose: further refinement based on neutron-diffraction data. *Acta Crystallogr. B* **35**, 656–659 (1979).
18. Brady, J. W. Molecular dynamics simulations of α -D-glucose. *J. Am. Chem. Soc.* **108**, 8153–8160 (1986).
19. Brady, J. W. Molecular dynamics simulations of β -D-glucopyranose. *Carbohydr. Res.* **165**, 306–312 (1987).
20. Joshi, N. V. & Rao, V. S. R. Flexibility of the pyranose ring in α - and β -D-glucoses. *Biopolymers* **18**, 2993–3004 (1979).
21. Bobbitt, J. M. in *Advances in Carbohydrate Chemistry* (eds Wolfram, M. L. & Tipson, R. S.) 1–4 (Academic, New York, 1956).
22. Smith, S. B., Finzi, L. & Bustamante, C. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science* **258**, 1122–1126 (1992).
23. Li, H., Rief, M., Oesterhelt, F. & Gaub, H. E. Single-molecular force spectroscopy on xanthan by AFM. *Adv. Mater.* **3**, 316–319 (1998).
24. Kellie, G. M. & Riddell, F. G. Non-chair conformations of six-membered rings. *Top. Stereochem.* **8**, 225–264 (1974).
25. Pickett, H. M. & Strauss, H. L. Conformational structure, energy, and inversion rates of cyclohexane and some related oxanes. *J. Am. Chem. Soc.* **92**, 7281–7290 (1970).
26. Drickamer, K. Making a fitting choice: common aspects of sugar-binding sites in plant and animal lectins. *Structure* **5**, 465–468 (1997).
27. Barton, D. H. R. The principles of conformational analysis. *Science* **169**, 539–544 (1970).
28. Florin, E. L. *et al.* Sensing specific molecular interactions with the atomic force microscope. *Biosensors Bioelectr.* **10**, 895–901 (1995).
29. Quanta97/CHARMM (Molecular Simulations, 9685 Scranton Rd, San Diego, California 92121, USA, 1997).
30. Gaussian 94 (Gaussian Inc., Carnegie Office Prk, Bldg 6, Pittsburgh, Pennsylvania 15106, USA, 1994).

Acknowledgements. This work was supported by the NSF (P.E.M.) and the NIH (J.M.F.).

Correspondence and requests for materials should be addressed to J.M.F. (e-mail: fernandez,julio@mayo.edu).

Effect of interannual climate variability on carbon storage in Amazonian ecosystems

Hanqin Tian*, Jerry M. Melillo*, David W. Kicklighter*, A. David McGuire†, John V. K. Helfrich III*, Berrien Moore III‡ & Charles J. Vörösmarty‡

* The Ecosystems Center, Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA

† US Geological Survey, Alaska Cooperative Fish and Wildlife Research Unit, University of Alaska, Fairbanks, Alaska 99775, USA

‡ Institute for the Study of Earth, Oceans and Space, University of New Hampshire, New Hampshire 03824, USA

The Amazon Basin contains almost one-half of the world's undisturbed tropical evergreen forest as well as large areas of tropical savanna^{1,2}. The forests account for about 10 per cent of the world's terrestrial primary productivity and for a similar fraction of the carbon stored in land ecosystems^{2,3}, and short-term field measurements⁴ suggest that these ecosystems are globally important carbon sinks. But tropical land ecosystems have experienced substantial interannual climate variability owing to frequent El Niño episodes in recent decades⁵. Of particular importance to climate change policy is how such climate variations, coupled with increases in atmospheric CO₂ concentration, affect terrestrial carbon storage^{6–8}. Previous model analyses have demonstrated the importance of temperature in controlling carbon storage^{9,10}. Here we use a transient process-based biogeochemical model of terrestrial ecosystems^{3,11} to investigate interannual variations of carbon storage in undisturbed Amazonian ecosystems in response to climate variability and increasing atmospheric CO₂ concentration during the period 1980 to 1994. In El Niño years, which bring hot, dry weather to much of the Amazon region, the ecosystems act as a source of carbon to the atmosphere (up to 0.2 petagrams of carbon in 1987 and 1992). In other years, these ecosystems act as a carbon sink (up to 0.7 Pg C in 1981 and

1993). These fluxes are large; they compare to a 0.3 Pg C per year source to the atmosphere associated with deforestation in the Amazon Basin in the early 1990s¹². Soil moisture, which is affected by both precipitation and temperature, and which affects both plant and soil processes, appears to be an important control on carbon storage.

Carbon fluxes calculated using the Terrestrial Ecosystem Model (TEM) include net primary production (NPP), microbial respiration (R_H) and net ecosystem production (NEP), all of which are influenced by climate. Net primary production, the net amount of carbon captured by plants, is also influenced by atmospheric CO₂ concentration. The difference between NPP and R_H equals NEP, which is equivalent to annual net carbon storage or loss for an ecosystem.

For regional or global extrapolations with TEM, we use input data on vegetation, elevation, soil texture, monthly mean temperature, monthly mean precipitation and monthly mean solar radiation. The input data sets are gridded at a resolution of 0.5° latitude by 0.5° longitude. The structure, parametrization, calibration and performance of TEM have been documented previously^{3,11}.

We first ran TEM in equilibrium mode to generate an initial condition for the transient runs, using the long-term mean of monthly temperature, monthly precipitation, monthly solar radiation and the level of atmospheric CO₂ concentration at the beginning of this century (296 p.p.m.y.). Then we ran TEM in transient mode using historical input data from 1900 to 1994 including: (1) historical mean atmospheric CO₂ concentration generated from atmospheric and ice core CO₂ observations¹³, and (2) historical monthly data for air temperature¹⁴ and precipitation¹⁵. The historical temperature and precipitation data were interpolated by the Max Planck Institute for Meteorology to a 0.5° spatial resolution. Two transient runs were made: one considering the climate and CO₂ transients together, and one considering only the climate transient. A comparison of these two runs was used to determine the effect of CO₂ 'fertilization' on carbon storage. All other model analyses were based on the combined climate and CO₂ transient run. We used the IGBP-DIS land-cover data set¹⁶ as a basis for adjusting the area of undisturbed ecosystems in the Basin to account for land-cover conversions such as forest to cropland. This reduces the area of undisturbed ecosystems by 11%, which is in good agreement with other land-cover change estimates for the Basin^{12,17}.

For the Amazon Basin, TEM results (Table 1) show that annual NEP varied from –0.2 Pg C (10¹⁵ g C) in 1987 and 1992, to 0.7 Pg C in 1981 and 1993, because of the combined effects of climate variability and increasing atmospheric CO₂ concentration. A negative NEP means these ecosystems are a source of atmospheric CO₂,

Table 1 Interannual variations in carbon fluxes in Amazonian ecosystems

Year	Climate with CO ₂			Climate only		
	NPP (Pg C yr ⁻¹)	R_H (Pg C yr ⁻¹)	NEP (Pg C yr ⁻¹)	NPP (Pg C yr ⁻¹)	R_H (Pg C yr ⁻¹)	NEP (Pg C yr ⁻¹)
1980	5.0	4.7	0.3	4.5	4.5	0.0
1981	5.4	4.7	0.7	4.9	4.4	0.5
1982	4.9	4.8	0.1	4.3	4.5	–0.2
1983	4.8	4.9	–0.1	4.2	4.6	–0.4
1984	5.3	4.8	0.5	4.7	4.5	0.2
1985	5.0	4.8	0.2	4.4	4.4	0.0
1986	5.2	4.8	0.4	4.7	4.5	0.2
1987	4.7	4.9	–0.2	4.1	4.6	–0.5
1988	5.0	4.9	0.1	4.3	4.5	–0.2
1989	5.1	4.8	0.3	4.6	4.5	0.1
1990	5.1	4.8	0.3	4.5	4.5	0.0
1991	4.8	4.9	–0.1	4.1	4.5	–0.4
1992	4.6	4.8	–0.2	3.8	4.4	–0.6
1993	5.7	5.0	0.7	5.1	4.5	0.6
1994	5.3	5.0	0.3	4.7	4.6	0.1
s.d.	0.3	0.1	0.3	0.3	0.1	0.3

The carbon fluxes between the atmosphere and undisturbed Amazonian ecosystems in response to interannual climate variability and increasing atmospheric CO₂ concentration as estimated by the Terrestrial Ecosystem Model. NPP, net primary production; R_H , heterotrophic respiration; NEP, net ecosystem production; s.d., standard deviation. A negative NEP indicates a net flux of carbon from the land to the atmosphere.

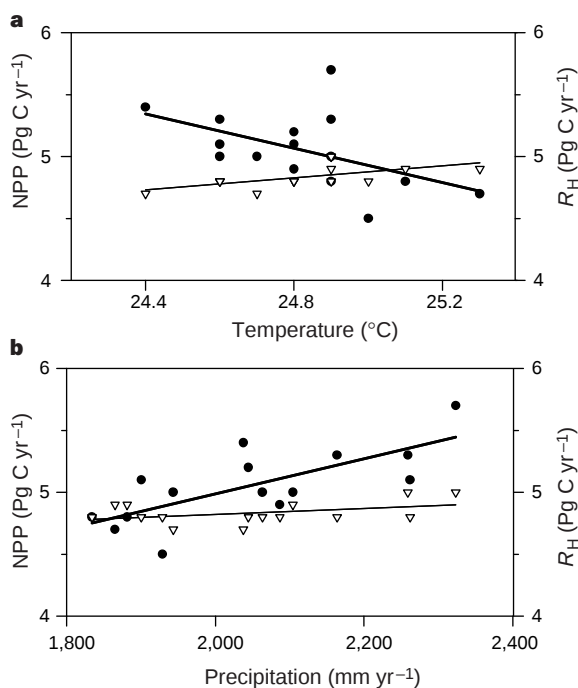


Figure 1 Relation of carbon fluxes to temperature and precipitation. **a**, Relations of annual net primary production (NPP, thick line, circles) and annual heterotrophic respiration (R_H , thin line, triangles) to annual mean temperature in the combined simulation of transient climate and transient atmospheric CO_2 . **b**, Relations of annual net primary production (NPP) and annual heterotrophic respiration (R_H) to annual precipitation in the combined simulation of transient climate and transient atmospheric CO_2 . Annual NPP is negatively correlated with annual mean temperature ($R = -0.51$, $P < 0.05$), and positively correlated with annual precipitation ($R = 0.73$, $P < 0.01$). Annual R_H is positively correlated with temperature ($R = 0.60$, $P < 0.01$), but not significantly correlated with precipitation.

whereas a positive NEP means that these ecosystems are a sink for atmospheric CO_2 .

Model results indicate that CO_2 'fertilization' of the vegetation of the Amazon increases NEP. Climate variability with CO_2 fertilization generally resulted in a higher annual NEP than did climate without CO_2 fertilization (Table 1). The strength of the CO_2 fertilization effect for the Amazon Basin was between 0.1 to 0.4 Pg C yr^{-1} for the 15-year period 1980–1994. The CO_2 effect includes both the direct stimulation of plant growth by CO_2 and the indirect enhancement of plant water use efficiency^{6,11}.

According to the TEM simulations, year-to-year variations in carbon storage are closely related to shifts among the phases of El Niño/Southern Oscillation. Undisturbed Amazonian ecosystems acted as a source of atmospheric CO_2 (that is, negative NEP) during El Niño years, 1982/1983, 1987/1988 and 1991/1992. An El Niño event results in drier and warmer weather conditions for the Amazon Basin¹⁸. Both drier weather and warmer temperatures decrease NPP, and warmer temperatures increase R_H (Fig. 1). In

other years, which are wetter and cooler, terrestrial ecosystems in the Amazon Basin acted as a sink for atmospheric CO_2 (that is, positive NEP).

For the period of this study, interannual variations of NEP are correlated with the interannual variations of NPP. Annual NPP shows much larger variations than annual R_H (Table 1). A major controller of NPP in the Basin appears to be soil moisture, which is a function of precipitation and temperature. Soil moisture represents water availability to plants; it controls, in part, the transformation of organically bound soil nitrogen to inorganic soil nitrogen, the nitrogen most readily available to plants^{19,20}. Our analysis shows that an increase in precipitation leads to increases in soil moisture ($R = 0.73$, $P < 0.002$) and net nitrogen mineralization ($R = 0.56$, $P < 0.03$). The model results, indicating a close link between soil moisture and nitrogen mineralization, are consistent with field measurements made in the western Amazon¹⁹.

Consistency between model results and field measurements is essential for establishing the credibility of a model such as TEM. The results of TEM are in reasonable agreement with measurement-based estimates of: (1) short-term, site-specific NEP; and (2) field-based estimates of basin-wide carbon stocks in vegetation and soils. At three sites in the Basin, two forests^{4,21} and one savanna²², the technique of eddy covariance has been used to estimate net carbon exchange between these ecosystems and the atmosphere. We ran TEM in site-specific mode for each place, using the climate for the period of the field study. The model-derived estimate was the same as the field estimate for the forest in Rondônia, western Amazon; 16% lower than the field estimate for the forest near Manaus in the central Amazon, and 22% lower than the field estimate for the savanna at the Reserva Ecológica de Águas Emendadas, Brazil (Table 2).

For the current climate and today's atmospheric CO_2 concentration, the TEM estimate of mean carbon density in the forest vegetation of the Basin is 14 kg C m^{-2} , which is within the range of 13.6 and 14.9 kg C m^{-2} estimated from field surveys^{23,24}. Likewise, the mean Basin soil organic carbon density of 9.3 kg C m^{-2} , estimated with TEM, is close to the estimate of 10.3 kg C m^{-2} based on the RADAM Brazil field survey²⁵.

Fan *et al.*²¹ have used their short-term, site-specific estimates of NEP to extrapolate to the forests of the whole Basin for 1987, and Grace *et al.*⁴ did the same for the period July 1992 to June 1993. Both groups^{4,21} estimated that the Basin's forests were acting as a major carbon sink: 1.2 Pg C yr^{-1} for 1987 and 0.5 Pg C yr^{-1} for the July 1992 to June 1993 period. In contrast, TEM simulations show the forests of the Basin to be a carbon source of -0.2 Pg C for 1987, and to be in balance for the period July 1992 to June 1993. The TEM results for the period July 1992 to June 1993 reflect the fact that the region was moving out of an El Niño in 1992, and into a neutral year in 1993. For the period July–December 1992, TEM indicated that the basin functioned as a strong carbon source (-0.6 Pg C); for the period January to June 1993, TEM showed that the Basin acted as a strong sink (0.6 Pg C).

The difference between the Basin-wide estimates of Fan *et al.*²¹ and Grace *et al.*⁴ and the ones from TEM is at least in part because both groups^{4,21} made the simplifying assumption that the vegeta-

Table 2 Modelled and field net ecosystem production estimates compared

Ecosystems/Location	Time period	Field-based estimate*	TEM-based estimate†	Refs
Tropical rain forests/Rondônia (10°05' S, 61°57' W)	May ~ June/93	$0.6 \text{ g C m}^{-2} \text{ d}^{-1}$	$0.6 \text{ g C m}^{-2} \text{ d}^{-1}$	4
Tropical rain forests/Manaus (2°56' S, 59°57' W)	April ~ May/87	$0.6 \text{ g C m}^{-2} \text{ d}^{-1}$	$0.5 \text{ g C m}^{-2} \text{ d}^{-1}$	21
Savanna/Emendadas (15°33' S, 47°36' W)	March ~ May/95	$1.8 \text{ g C m}^{-2} \text{ d}^{-1}$	$1.4 \text{ g C m}^{-2} \text{ d}^{-1}$	22

The field-based estimates for the two forest sites were average rates of net carbon exchange for the measurement periods. The field-based estimate for the savanna site was the maximum rate, the only rate reported by Miranda *et al.*²².

* The three field-based estimates of net ecosystem production were measured by the eddy covariance technique.

† We used gridded historical climate data as inputs to TEM for Rondônia and Emendadas. For Manaus, we use climate data from the site's weather station as inputs to TEM.

tion, soils and climate are uniform across the Basin for the period of interest. In contrast, the TEM estimates account for spatial variability of vegetation, soils and climate that give rise to place to place differences in NEP (Fig. 2). During El Niño years, annual NEP is negative at most sites in the Amazon Basin, but it does remain positive in some parts of the Basin, including the northwest corner. During non-El Niño years, most parts of the Basin have a positive annual NEP. Thus, site-specific NEP measurements made in the

field must be used with caution when extrapolating over space and time to develop Basin-wide estimates.

Several global-scale analyses have been undertaken to explore the relation between climate variability and carbon cycling. Kindermann *et al.*⁹ used a physiologically based model of the carbon budget in terrestrial ecosystems, the Frankfurt Biosphere Model (FBM), for the period 1980–1993; their general conclusions with respect to the tropical regions, defined as the latitudinal band from 30°N to 30°S, were that annual NEP of land ecosystems in the tropics is particularly sensitive to interannual variations in climate, and that interannual temperature variations cause the main effects on annual NEP, with precipitation variations being important in some places.

Although our analysis of the TEM results also indicates that NEP in tropical ecosystems of the Amazon Basin is sensitive to inter-annual climate variability, we suggest that changes in the amount and the timing of precipitation cause the main effects on annual NEP in this region. Our analysis further indicates that changes in precipitation combine with changes in temperature to affect soil moisture, the factor we have identified as an important controller of carbon storage in the Amazon Basin. This conclusion agrees well with field results from a forest near Manaus, Brazil²⁶.

With reference to the consequences for NEP of precipitation changes in the tropics, Kindermann *et al.*⁹ indicated that the drier tropical systems, including the drought deciduous forests, are affected most. Analysis of the TEM data for the major vegetation types in the Basin indicates that whereas changes in precipitation resulted in the largest relative changes in net carbon storage per unit area in the dry ecosystems, the largest absolute changes in net carbon storage per unit area occurred in the moist and wet forests of the Basin.

Braswell *et al.*¹⁰ investigated how interannual temperature variability affects net carbon storage in terrestrial ecosystems at the global scale and found a significant relationship between atmospheric CO₂ growth rate and temperature. Their analysis suggested that the terrestrial response to changes in temperature results in either enhanced plant production, reduced heterotrophic respiration, or both, such that global NEP is positive about two years after an El Niño event.

Our results for the Amazon region differ from the global analysis of Braswell *et al.*¹⁰. Although the TEM simulations indicate that both NPP and NEP are significantly correlated with 'current' air temperature and precipitation, our analyses indicate no significant two-year lag effect between temperature and NEP for undisturbed ecosystems in the Amazon Basin. This result is consistent with the fact that year-to-year temperature variations in the Basin are small. Combining the Braswell *et al.* analysis¹⁰ with ours, we infer that the two-year lag phenomenon must be occurring in the extra-tropical areas where El Niño events cause large increases in temperature⁵.

An additional issue is the long-term effect of interannual climate variability on regional and global carbon storage. Are carbon losses in some years balanced by carbon gains in others? Or is there any reason to believe that land ecosystems are on longer-term trajectories of carbon gain or loss? The TEM simulations suggest that during the fifteen-year period 1980–1994, the undisturbed ecosystems of the Amazon Basin accumulated a total of ~3.3 Pg C, or an average of ~0.2 Pg C yr⁻¹. The progressive increase in atmospheric CO₂ concentration is the factor most likely to be responsible for the accumulation (Table 1).

An average net carbon storage of 0.2 Pg C yr⁻¹ in the Basin has both regional and global implications. At the regional scale, a net storage of 0.2 Pg C yr⁻¹ is the same order of magnitude as the estimates of the net annual release of carbon to the atmosphere from deforestation in this region during the period from the 1970s to the 1990s^{12,27}. For the 1970s, Skole²⁷ estimated the release of carbon from the Brazilian Amazon to be 0.1 Pg C yr⁻¹ due to deforestation. For the early 1990s, Fearnside¹² estimated the defor-

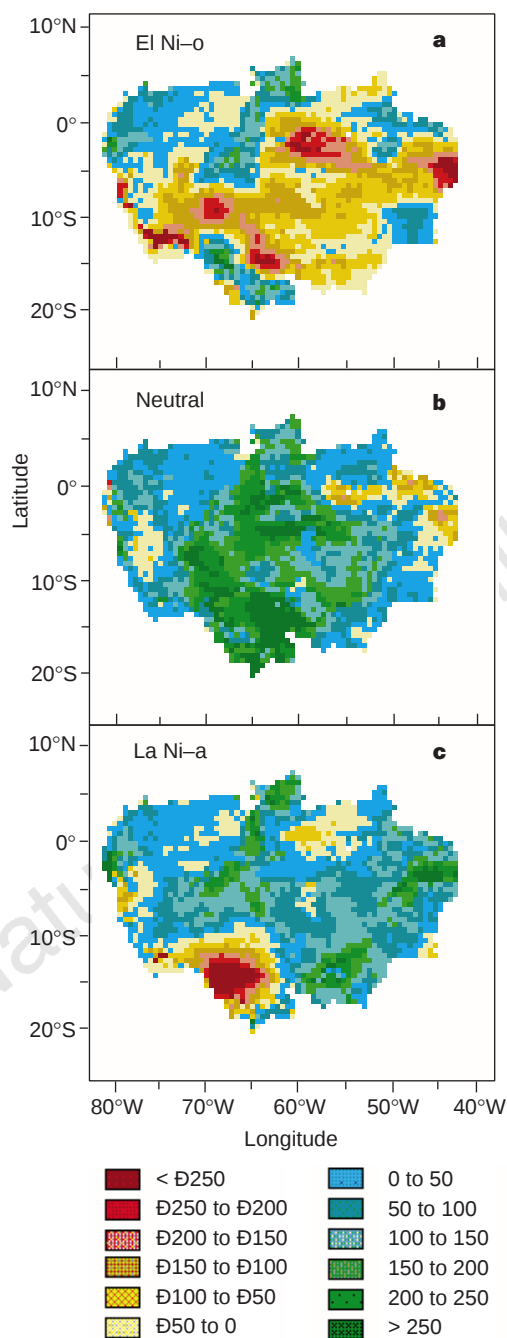


Figure 2 Net ecosystem production across the Amazon Basin. Spatial variability in net ecosystem production (g C m⁻² yr⁻¹) in the combined simulation of transient climate and transient atmospheric CO₂ during three phases of El Niño/Southern oscillation: **a**, an El Niño year (1987); **b**, a neutral year (1981); and **c**, a La Niña year (1989). Regions that act as a source of atmospheric carbon (annual NEP is negative) are designated by shades of brown, red, or yellow, and regions that act as a sink of atmospheric carbon (annual NEP is positive) are designated by shades of blue or green.

estation-caused release of carbon from the Brazilian Amazon rose to 0.3 Pg. Thus, the TEM simulation suggests that releases of CO₂ from tropical deforestation in the Basin were partly balanced by CO₂ uptake by undisturbed ecosystems.

According to a recent IPCC analysis of the global carbon budget for the 1980s, terrestrial ecosystems were acting as a net sink for ~1.8 PgC per year⁷. The TEM simulation for the 1980s indicates that the undisturbed ecosystems in the Amazon Basin may have accounted for 13% of this terrestrial carbon sink. These decadal averages, of course, mask the effects of interannual climate variability on NEP in the world's terrestrial ecosystems. One of the next tasks is to check the results of global-scale analyses against measurements of well constrained global indexes of land-atmosphere carbon exchanges, such as the ratio of atmospheric N₂ to O₂ (ref. 28). Regional- and global-scale information on the spatial and temporal patterns of carbon uptake and release by terrestrial ecosystems, and the biogeochemical processes responsible for these patterns, are needed as we enter a post-Kyoto period in which we aim to manage the global carbon cycle.

Received 5 June; accepted 18 September 1998.

- Melillo, J. M., Houghton, R. A., Kicklighter, D. W. & McGuire, A. D. Tropical deforestation and the global carbon budget. *Annu. Rev. Energy Environ.* **21**, 293–310 (1996).
- Keller, M., Melillo, J. M. & de Mello, W. Z. Trace gas emissions from ecosystems of the Amazon basin. *Cienc. Cult. J. Braz. Assoc. Adv. Sci.* **49**, 87–97 (1997).
- Melillo, J. M. *et al.* Global climate change and terrestrial net primary production. *Nature* **363**, 234–240 (1993).
- Grace, J. *et al.* Carbon dioxide uptake by an undisturbed tropical rain forest in southwest Amazonia, 1992 to 1993. *Science* **270**, 778–780 (1995).
- Nicholls, N. *et al.* in *Climate Change 1995: The Science of Climate Change* (eds Houghton, J. T. *et al.*) 132–192 (Cambridge University Press, Cambridge, UK, 1996).
- Melillo, J. M., Prentice, I. C., Farquhar, G. D., Schulze, E.-D. & Sala, O. E. in *Climate Change 1995: The Science of Climate Change* (eds Houghton, J. T. *et al.*) 444–481 (Cambridge University Press, Cambridge, UK, 1996).
- Schimel, D. S. *et al.* in *Climate Change 1995: The Science of Climate Change* (eds Houghton, J. T. *et al.*) 67–131 (Cambridge University Press, Cambridge, UK, 1996).
- Dai, A. & Fung, I. Y. Can climate variability contribute to the “missing” CO₂ sink? *Glob. Biogeochem. Cyc.* **7**, 599–609 (1993).
- Kindermann, J., Wurth, G., Kohlmaier, G. H. & Badeck, E.-W. Interannual variation of carbon exchange fluxes in terrestrial ecosystems. *Glob. Biogeochem. Cyc.* **10**, 737–755 (1996).
- Braswell, B. H., Schimel, D. S., Linder, E. & Moore III, B. The response of global terrestrial ecosystems to interannual temperature variability. *Science* **278**, 870–872 (1997).
- Tian, H., Melillo, J. M., Kicklighter, D. W., McGuire, A. D. & Helfrich, J. The sensitivity of terrestrial carbon storage to historical atmospheric CO₂ and climate variability in the United States. *Tellus* (in the press).
- Fearnside, P. M. Greenhouse gases from deforestation in Brazilian Amazonia: net committed emission. *Climatic Change* **35**, 321–360 (1997).
- Enting, I. G., Wigley, T. M. L. & Heimann, M. *CSIRO Div. Atmos. Res. Techn. Pap. No.* **31**, 1–120 (1994).
- Jones, P. D. Hemispheric surface air temperature variations: a reanalysis and an update to 1993. *J. Clim.* **7**, 1794–1802 (1994).
- Hulme, M. A *Historical Monthly Precipitation Data Set for Global Land Area from 1900 to 1994*. Univ. East Anglia, Norwich, UK (1995).
- Loveland, T. R. & Belward, A. S. The IGBP-DIS 1 km land-cover data set, DISCOVER—first results. *Int. J. Remote Sensing* **18**, 3291–3295 (1997).
- Skole, D. & Tucker, C. J. Tropical deforestation and habitat fragmentation in the Amazon: satellite data from 1978 to 1988. *Science* **260**, 1905–1910 (1993).
- Vörösmarty, C. J. *et al.* Analyzing the discharge regime of a large tropical river through remote sensing, ground-based climate data, and modeling. *Water Resources Res.* **32**, 3137–3150 (1996).
- Neill, C. *et al.* Nitrogen dynamics in soils of forests and active pastures in the western Brazilian Amazon Basin. *Soil Biol. Biochem.* **27**, 1167–1175 (1995).
- Melillo, J. M., Kicklighter, D., McGuire, A., Peterjohn, W. & Newkirk, K. in *Dahlem Conf. Proc.* 175–189 (Wiley, New York, 1995).
- Fan, S. M., Wofsy, S. C., Bakwin, P. S. & Jacob, D. J. Atmosphere–biosphere exchange of CO₂ and O₃ in the central Amazon forest. *J. Geophys. Res.* **95**, 16851–16864 (1990).
- Miranda, A. C. *et al.* Fluxes of carbon, water and energy over Brazilian cerrado: an analysis using eddy covariance and stable isotopes. *Plant. Cell Environ.* **20**, 315–328 (1997).
- Brown, S. A. & Lugo, A. E. Aboveground biomass estimates for tropical moist forests of the Brazilian Amazon. *Interciencia* **17**, 8–18 (1992).
- Fearnside, P. M. Forest biomass in Brazilian Amazonia: comments on the estimate by Brown and Lugo. *Interciencia* **17**, 19–27 (1992).
- Moraes, J. F. L. *et al.* Soil carbon stocks of the Brazilian Amazon Basin. *Soil Sci. Soc. Am. J.* **59**, 244–247 (1995).
- Williams, M. *et al.* Seasonal variation in net carbon exchange and evapotranspiration in a Brazilian rain forest: A modelling analysis. *Plant, Cell Environ.* (in the press).
- Skole, D. *Measurement of Deforestation in the Brazilian Amazon Using Satellite Remote Sensing*. Thesis, Univ. New Hampshire (1992).
- Keeling, R. F., Piper, S. C. & Heimann, M. Global and hemispheric CO₂ sinks deduced from changes in atmospheric O₂ concentration. *Nature* **381**, 218–221 (1996).

Acknowledgements. This work was supported by the Earth Observing System Program of the National Aeronautics and Space Administration and the Electric Power Research Institute through the Carbon Cycle Model Linkage Project (CCMLP) and the Vegetation/Ecosystem Modeling and Analysis Project (VEMAP). We thank M. Heimann and his coworkers at the Max Planck Institute, Hamburg, for climatic and CO₂ data, and C. Prentice, S. Brown, R. Huang, C. Neill and M. Williams for their critical comments.

Correspondence and requests for materials should be addressed to H.T. (e-mail: htian@lupine.mbl.edu).

Ratios of ferrous to ferric iron from nanometre-sized areas in minerals

Laurence A. J. Garvie* & Peter R. Buseck**

* Department of Geology, † Department of Chemistry/Biochemistry, Arizona State University, Tempe, Arizona 85287-1404, USA

Minerals with mixed valence states are widespread and form in many different rock types¹. They can contain, for example, Fe²⁺–Fe³⁺ and Mn²⁺–Mn³⁺–Mn⁴⁺, with the ratios of oxidation states reflecting the redox conditions under which the host materials crystallized. The distribution of the ratio of iron (III) to total iron content (Fe³⁺/ΣFe) in minerals reflects the oxidation states of their host rocks and is therefore important for answering fundamental questions about the Earth's evolution and structure^{2–8}. Iron is the most sensitive and abundant indicator of oxidation state, but many mineral samples are too fine-grained and heterogeneous to be studied by standard methods such as Mössbauer spectroscopy, electron microprobe, and wet chemistry. Here we report on the use of electron energy-loss spectroscopy with a transmission electron microscope to determine Fe³⁺/ΣFe in minerals at the nanometre scale. This procedure is efficient for determining Fe³⁺/ΣFe ratios of minor and major amounts of iron on a scale heretofore impossible and allows information to be obtained not only from ultra-fine grains but also, for example, at reaction fronts in minerals.

Iron is abundant and occurs in oxidation states that range from Fe⁰ (for example, at the core–mantle boundary) to Fe³⁺ (at the Earth's surface). Estimates of the oxidation state of the Earth's upper mantle have been controversial, at least in part because of differences in the results produced by the various methods used to determine Fe³⁺/ΣFe values⁹. Mössbauer spectroscopy and wet chemical methods are bulk techniques, and do not provide information on fine-grained minerals or fine-scale heterogeneity. Micro-Mössbauer⁷, electron microprobe¹⁰, X-ray absorption^{11,12}, and X-ray photoelectron^{13,14} spectroscopies optimistically allow regions with diameters from 1 to 50 μm to be analysed. However, fine-grained heterogeneous samples and zoning at the sub-micrometre level

Table 1 Comparison of EELS and published Fe³⁺/ΣFe ratios

Sample	EELS	Published data		
	Fe ³⁺ /ΣFe	Fe ³⁺ /ΣFe	FeO _{tot}	Refs
Amphibole LC3*	0.75	0.76	11.08	28
Amphibole SC1*	0.52	0.49	16.69	28
Amphibole SC3*	0.31	0.26	16.83	28
Amphibole LC1*	0.65	0.63	10.96	28
Amphibole Kaert	0.93	0.93	11.82	29
Augite CVF1*	0.38	0.44	8.01	28
Augite PX4*	0.28	0.25	6.83	28
Augite SC4*	0.33	0.36	8.12	28
Augite SC6*	0.29	0.31	13.45	28
Augite LC4*	0.40	0.50	6.99	28
Glass 7‡	0.17	0.12	11.46	30
Glass 5‡	0.33	0.25	11.46	30
Glass air‡	0.82	0.70	11.46	30
Spinel KR35§	0.21	0.23 (0.29)	11.87	5

* Amphiboles LC1, LC3, SC1, SC3 and augites are from megacrysts in alkali basalts; Fe³⁺/ΣFe is from wet chemistry.

† Amphibole Kaert, Kaersutite from USNM no. 116503.0032. Fe³⁺/ΣFe is from Mössbauer spectroscopy.

‡ Glasses prepared from 1921 Kilauea basalt with CO/CO₂ gases at log *f*_{O₂} of –8, –7, –5 and in air at 1 atm and 1,400 °C; Fe³⁺/ΣFe is estimated from the activity–composition relationships in the system Fe–Pt.

§ Spinel from a mantle xenolith; Fe³⁺/ΣFe is from Mössbauer spectroscopy and, in brackets, microprobe data using the Bence–Albee data reduction method.

cannot be analysed using these methods, and so $\text{Fe}^{3+}/\Sigma\text{Fe}$ determined from such materials can give erroneous results.

Geological samples are commonly zoned, and many experimental run products are too small or fine-grained for electron microprobe analysis. For example, lower-mantle (transition zone and below) mineral assemblages are particularly difficult to study because they are not normally present at the Earth's surface except as inclusions in diamonds of lower-mantle origin^{7,15}. In addition, information on lower-mantle conditions is inferred from theory and from multi-anvil and diamond-cell experiments^{16,17}, which indicate a predominance of (Mg,Fe)SiO₃ perovskite and (Mg,Fe)O magnesiowüstite in the upper part of the lower mantle. A pressing question from these experiments is the oxidation state of Fe in the perovskite¹⁷. However, the products from such diamond-cell experiments occur in minute amounts, and require measurements made with exceptionally high spatial resolution.

The spatial-resolution problem when determining $\text{Fe}^{3+}/\Sigma\text{Fe}$ can be overcome by using electron energy-loss spectroscopy (EELS) combined with a transmission electron microscope (TEM). EELS is valuable for determining elemental composition, coordination, oxidation state and spin state^{18–20}. Additionally, diffraction patterns and energy-dispersive X-ray spectra can be recorded from the same areas, complementing the EELS data. Several EELS studies show that the 3d transition-metal L_{2,3} edges have spectral shapes that are sensitive to the oxidation state of the metal^{19,21}. An EELS spectrum displays electron intensity as a function of energy loss²². Core-loss edges result from the transition of core electrons to unoccupied states in the conduction band. Core-loss-edge onsets represent the ionization thresholds—the energies that approximately correspond to the inner-shell binding energies—and are characteristic of the elements. The Fe L_{2,3} edges are characterized by sharp maxima just above the edge onsets, and represent transitions to empty d states.

We used a Gatan parallel EELS with a TEM to acquire the Fe L_{2,3} spectra. Fourteen single-phase materials containing mixed-valence

iron were studied, covering a wide range of $\text{Fe}^{3+}/\Sigma\text{Fe}$ values (Table 1). The total iron (FeO_{tot}) values are in the range of important crustal and mantle minerals. EELS spectra of Fe^{2+} - and Fe^{3+} -bearing minerals show distinct L_{2,3} edge shapes and chemical shifts (Fig. 1). The Fe^{2+} L₃ edge has a sharp maximum at ~707.5 eV, whereas Fe^{3+} L₃ edges show maxima at ~708.0 and 709.5 eV. The pre-peak to the main Fe^{3+} L₃ peak occurs at the same energy as the main Fe^{2+} L₃ peak, but *ab initio* calculations²³ show it to be from Fe^{3+} . These calculations also show that the structure at ~710 eV on the fayalite L₃ edge arises from Fe^{2+} .

Spectra from a range of minerals show a constancy of edge shapes for given oxidation states (Fig. 2). This constancy demonstrates the feasibility of determining $\text{Fe}^{3+}/\Sigma\text{Fe}$ values. In addition, metallic iron has an L_{2,3} edge shape different from that of Fe^{2+} and Fe^{3+} , with broad asymmetrical L₃ and L₂ peaks. Similarly, the distinct L_{2,3}-edge shapes of Fe^0 and Fe^{n+} allow the possibility of measuring the $\text{Fe}^0/\text{Fe}^{n+}$ ratio. The crystallinity of the sample has little effect on the Fe L_{2,3} shape because spectra of Fe^{2+} and Fe^{3+} are dominated by quasi-atomic transitions, with only minor modification from the solid-state environment^{19,23}.

Minerals containing Fe^{2+} and Fe^{3+} show two-peaked L₃ edges that are intermediate in shape between the two single-valence end-members (Fig. 3). The changes in relative heights of the L₃ peaks qualitatively reflect the published $\text{Fe}^{3+}/\Sigma\text{Fe}$ values. The mixed-

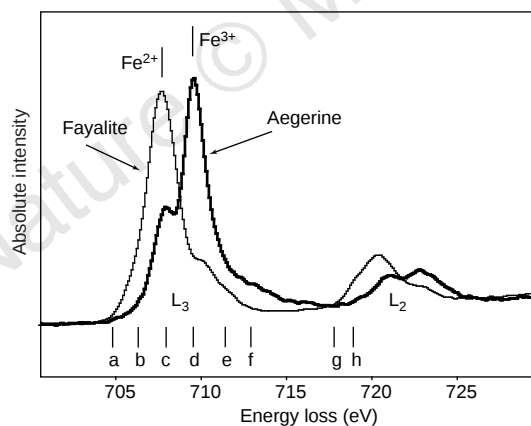


Figure 1 Comparison of the Fe L_{2,3} edges of fayalite and aegerine. The weaker L₂ edge rests on the tail of the more intense L₃ edge. The two spectra were scaled to each other by normalizing the edges to the continua above the edges, so the spectral intensities reflect the ratio of Fe^{2+} and Fe^{3+} d-holes. a, Fe^{2+} L₃ edge onset at 704.6 eV; b, Fe^{3+} L₃ edge onset at 705.9 eV; c, 708.5 eV; d, 709.5 eV; e, 711.6 eV; f, 713.1 eV; g, 717.8 eV; and h, 718.8 eV. Spectra were acquired with a Gatan 666 parallel EELS spectrometer attached to a Philips 400-ST field-emission gun TEM operated at an accelerating voltage of 100 keV. The energy resolution at the zero-loss peak was 0.7 eV. The spectra were acquired with convergence and collection angles of 16 and 11 mrad, respectively, a probe current of 10–15 nA, and acquisition times of 1–3 s per spectrum. For each sample, a small fraction of an ~100-μm crystal was deposited onto a TEM grid, and spectra were acquired from thin regions of ~20–100-nm diameter. Care was taken to ensure that samples were not damaged by the electron beam; such damage was a problem with the amphiboles, but cooling to liquid-nitrogen temperature slowed the damage process.

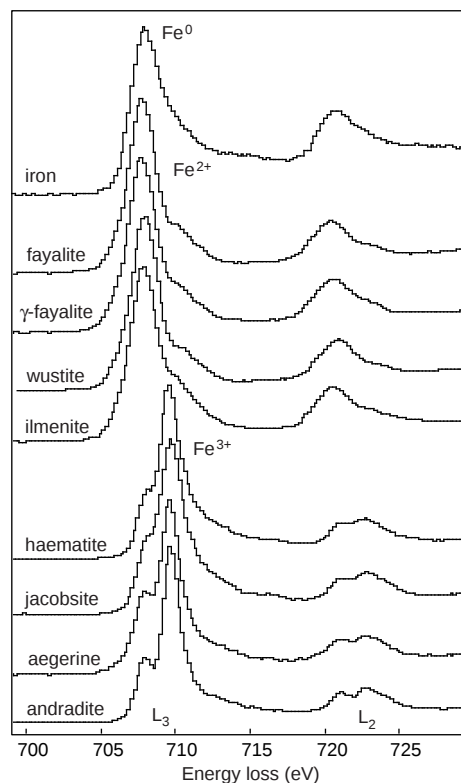


Figure 2 Fe L_{2,3} edges of elemental Fe and selected minerals containing single-valent Fe. The spectra from a range of minerals, with Fe bonded to O, show a constancy of edge shapes for given oxidation states. These minerals contain octahedrally coordinated Fe, but Fe L_{2,3} edges with other coordinations exhibit the same basic edge shapes at the energy resolution used. Crystallinity and crystal chemistry have little effect on the Fe^{2+} L_{2,3} shapes, and only minor effects on the Fe^{3+} -edge shapes. At higher energy resolutions minor differences resulting from site symmetry and crystal-field effects become apparent. The crystal field gives rise to peak broadening and gradual appearances of new peaks with increasing crystal-field interaction. Variations in energy of the Fe^{2+} and Fe^{3+} L₃ peak maxima among different minerals are less than ±1 eV. Iron, Fe; fayalite, Fe_2SiO_4 ; γ-fayalite, Fe_2SiO_4 ; wüstite, Fe_{1-x}O ; haematite, Fe_2O_3 ; jacobsonite, MnFe_2O_4 ; aegerine, $\text{NaFeSi}_2\text{O}_6$; andradite, $\text{Ca}_3\text{Fe}_2(\text{SiO}_4)_3$.

valence spectra were simulated by a multiple, linear, least-squares fitting method, with fayalite (Fe_2SiO_4) and aegerine ($\text{NaFeSi}_2\text{O}_6$) as the respective Fe^{2+} and Fe^{3+} model endmembers. The fayalite and aegerine spectra were scaled to each other by normalizing their $\text{L}_{2,3}$ peaks to the continua above the edges. In this way the integrated areas of the $\text{L}_{2,3}$ peaks are quantitatively related, and reflect the ratio of d -band holes of Fe^{2+} and Fe^{3+} . Before applying the fitting routine, the background and continuum beneath the L_3 and L_2 peaks were subtracted from all spectra. The d -like component of the spectra, representing the L_3 and L_2 maxima, was subtracted from the continuum part of the spectrum by using a straight line, from 705 to 730 eV, to separate the L_3 and L_2 peaks from the continuum.

Using the normalized fayalite and aegerine $\text{Fe L}_{2,3}$ edges, we were

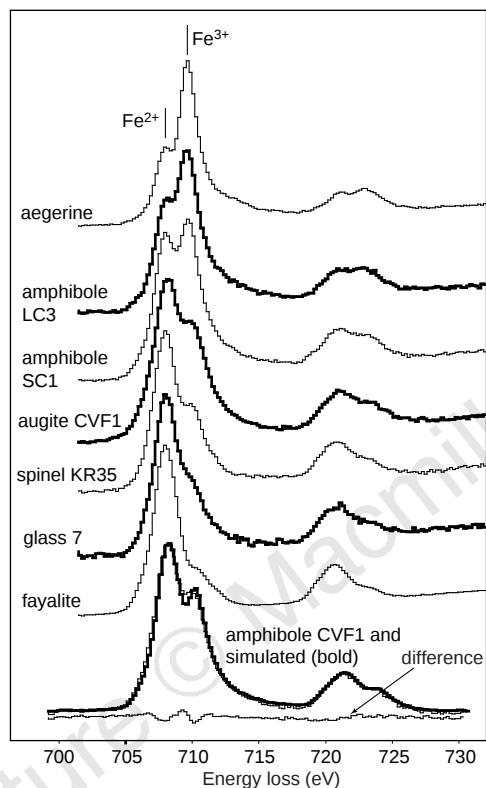


Figure 3 $\text{Fe L}_{2,3}$ edges of selected materials containing mixed-valent Fe. For comparison, two single-valence minerals, aegerine (Fe^{3+}) and fayalite (Fe^{2+}), are shown. The spectra are aligned with decreasing proportions of Fe^{3+} from top to bottom. At the bottom of the figure, the experimental spectrum, the corresponding fit using the model endmembers (bold line), and the differences between the two spectra are shown for amphibole CVF1. Aegerine, $\text{NaFeSi}_2\text{O}_6$; fayalite, Fe_2SiO_4 ; (see refs 28, 29, 30 and 5 for the compositions of the amphiboles, augites, spinel and glass respectively).

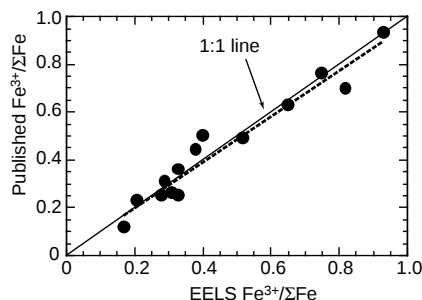


Figure 4 Comparison of $\text{Fe}^{3+}/\Sigma\text{Fe}$ from EELS and published results (from Table 1). The dashed line is the linear best fit for the data and is close to the 1:1 correlation line (solid line). The coefficient of determination, r^2 , is 0.95.

able to obtain good fits to the spectra from minerals having mixed oxidation states. The best-fit simulated spectra closely match the experimental peak shapes. Thus, minerals with Fe^{2+} and Fe^{3+} show spectra that are linear combinations of the separate Fe^{2+} and Fe^{3+} spectra, thereby allowing $\text{Fe}^{3+}/\Sigma\text{Fe}$ to be determined. The good agreement between the published $\text{Fe}^{3+}/\Sigma\text{Fe}$ values and those derived from our EELS measurements validates the latter (Table 1, Fig. 4). The range of samples illustrates the applicability of this method to many rock-forming minerals, provided that oxygen is the nearest-neighbour anion. The fits to the amphibole spectra were good despite the different octahedral crystallographic sites occupied by Fe.

Errors are difficult to estimate with the limited number of samples studied here but include, for example, energy resolution of the EELS-TEM, signal-to-noise ratio of the spectra, possible beam damage, the type of model spectra used for fitting, and the fitting routine. Except for the air-heated glass (Table 1), variations of $\text{Fe}^{3+}/\Sigma\text{Fe}$ between spectra from different parts of a single grain were less than $\pm 10\%$. Spectra from the air-heated glass gave $\text{Fe}^{3+}/\Sigma\text{Fe}$ ranging from 0.6 to 0.95, with the average of 14 spectra giving 0.82. The data (Table 1) can be fitted with a straight line with a coefficient of determination, r^2 , of 0.95. The least-squares fit is close to the 1:1 correlation line (Fig. 4).

Electron energy-loss spectroscopy with a TEM offers a unique combination of high-spatial with high-energy resolution^{24–26} and removes the fundamental spatial limitation imposed by most microbeam analytical techniques. This method complements the synchrotron micro-XANES method that allows the determination of $\text{Fe}^{3+}/\Sigma\text{Fe}$ at the micrometre scale¹¹. The spatial resolution is governed by the probe size in the TEM, which can be as small as 3 Å in a dedicated scanning TEM²⁴, although the actual probe size used is dictated by the electron-beam sensitivity of the sample and counting statistics. There is also the possibility of mapping Fe^{2+} – Fe^{3+} distributions with the TEM, using the EELS signal with a technique called electron spectroscopic imaging²⁷. Our study illustrates the applicability of EELS for determining $\text{Fe}^{3+}/\Sigma\text{Fe}$ by analysing well-characterized samples; this technique could be extended to other mixed-valence elements in minerals.

Note added in proof: While this Letter was in press, we were made aware of recently published work³¹ in which the $\text{Fe}^{3+}/\Sigma\text{Fe}$ ratio was determined using EELS with a TEM. Their method differs from ours and uses the differences in $\text{Fe L}_{2,3}$ white-line intensity ratios of Fe^{2+} and Fe^{3+} to determine $\text{Fe}^{3+}/\Sigma\text{Fe}$, rather than the fine structure that we use. □

Received 23 April; accepted 9 October 1998.

- Burns, R. G. in *Mixed Valency Systems: Applications in Chemistry, Physics and Biology* (ed. Prassides, K.) 175–199 (Kluwer Academic, Boston, 1991).
- Arculus, R. J. Oxidation status of the mantle: Past and present. *Annu. Rev. Earth Planet. Sci.* **13**, 75–95 (1985).
- Carmichael, I. S. E. & Ghiorso, M. S. Oxidation–reduction relations in basic magma: a case for homogenous equilibria. *Earth Planet. Sci. Lett.* **78**, 200–210 (1986).
- Wood, B. J. & Virgo, D. Upper mantle oxidation state: Ferric iron contents of ilherzolite spinels by ^{57}Fe Mössbauer spectroscopy and resultant oxygen fugacities. *Geochim. Cosmochim. Acta* **53**, 1277–1291 (1989).
- Canil, D., Virgo, D. & Scarfe, C. M. Oxidation state of mantle xenoliths from British Columbia, Canada. *Contrib. Mineral. Petrol.* **104**, 453–462 (1990).
- Luth, R. W., Virgo, D., Boyd, F. R. & Wood, B. J. Ferric iron in mantle derived garnets: Implications for thermobarometry and for the oxidation state of the mantle. *Contrib. Mineral. Petrol.* **104**, 56–72 (1990).
- McCammon, C., Hutchison, M. & Harris, J. Ferric iron content of mineral inclusions in diamonds from Sao Luiz: A view into the lower mantle. *Science* **278**, 434–436 (1997).
- Holland, H. D. Model for the evolution of the earth's atmosphere. *Geol. Soc. Am. (Budinington vol.)* 447–477 (1962).
- Virgo, D., Luth, R. W., Moats, M. A. & Ulmer, G. C. Constraints on the oxidation state of the mantle: An electrochemical and ^{57}Fe Mössbauer study of mantle-derived ilmenites. *Geochim. Cosmochim. Acta* **52**, 1781–1794 (1988).
- Hofer, H. E., Brey, G. P., Schulz-Dobrick, B. & Oberhänsli, R. The determination of the oxidation state of iron by the electron microprobe. *Eur. J. Mineral.* **6**, 407–418 (1994).
- Delaney, J. S., Bajt, S., Sutton, S. R. & Dyar, M. D. in *Mineral Spectroscopy: A Tribute to Roger G. Burns* (eds Dyar, M. D., McCammon, C. & Schaefer, M. W.) 165–171 (Spec. Publ. No. 5, Geochemical Soc., 1996).
- Droubay, T., Mursky, G. & Tonner, B. P. High-resolution x-ray absorption microspectroscopy of lamellar phases in natural ilmenite. *J. Electron. Spectrosc.* **84**, 159–169 (1997).
- Raeburn, S. P., Ilton, E. S. & Veblen, D. R. Quantitative determination of the oxidation state of iron in biotite using X-ray photoelectron spectroscopy: I. Calibration. *Geochim. Cosmochim. Acta* **61**, 4519–4530 (1997).

14. Raeburn, S. P., Ilton, E. S. & Veblen, D. R. Quantitative determination of the oxidation state of iron in biotite using X-ray photoelectron spectroscopy: II. In situ analyses. *Geochim. Cosmochim. Acta* **61**, 4531–4537 (1997).
15. Harris, J., Hutchison, M. T., Hursthouse, M., Light, M. & Harte, B. A new tetragonal silicate mineral occurring as inclusions in lower-mantle diamonds. *Nature* **387**, 486–488 (1997).
16. Mao, H.-K., Shen, G. & Hemley, R. J. Multivariable dependence of Fe-Mg partitioning in the lower mantle. *Science* **278**, 2098–2100 (1997).
17. Wood, B. J. & Rubie, D. C. The effect of alumina on phase transformations at the 660-kilometer discontinuity from Fe-Mg partitioning experiments. *Science* **273**, 1522–1524 (1996).
18. Disko, M. M., Ahn, C. C. & Fultz, B. *Transmission Electron Energy Loss Spectrometry in Materials Science* (Minerals, Metals & Materials Soc., Warrendale, IL, 1992).
19. Garvie, L. A. J. & Craven, A. J. High-resolution parallel electron energy-loss spectroscopy of Mn $L_{2,3}$ -edges in inorganic manganese compounds. *Phys. Chem. Minerals* **21**, 191–206 (1994).
20. Garvie, L. A. J. & Buseck, P. R. in *Boron Mineralogy, Petrology and Geochemistry* (eds Grew, E. S. & Anovitz, L. M.) 821–843 (*Reviews in Mineralogy* vol. 33, Mineralogical Soc. Am., Washington DC, 1996).
21. Garvie, L. A. J., Craven, A. J. & Brydson, R. Use of electron-energy loss near-edge fine structure in the study of minerals. *Am. Mineral.* **79**, 411–425 (1994).
22. Egerton, R. F. *Electron Energy-Loss Spectroscopy in the Electron Microscope* 2nd edn (Plenum, New York, 1996).
23. van der Laan, G. & Kirkman, I. W. The 2p absorption spectra of 3d transition metal compounds in tetrahedral and octahedral symmetry. *J. Phys.: Condens. Matter* **4**, 4189–4204 (1992).
24. Browning, N. D., Chisholm, M. F. & Pennycook, S. J. Atomic-resolution chemical analysis using a scanning transmission electron microscope. *Nature* **366**, 143–146 (1993).
25. Bradley, J. P. Nanometer-scale mineralogy and petrography of fine-grained aggregates in anhydrous interplanetary dust particles. *Geochim. Cosmochim. Acta* **58**, 2123–2134 (1994).
26. Colliex, C. New trends in STEM-based nano-EELS analysis. *J. Electron Microsc.* **45**, 44–50 (1996).
27. Czank, M., Mayer, J. & Klein, U. Electron spectroscopic imaging (ESI): A new method to reveal the existence of nm-scale exsolution lamellae. *Eur. J. Mineral.* **9**, 1199–1206 (1997).
28. Righter, K. & Carmichael, I. S. E. Mega-xenocrysts in alkali olivine basalts: Fragments of disrupted assemblages. *Am. Mineral.* **78**, 1230–1245 (1993).
29. McGuire, A. V., Francis, C. A. & Dyar, M. D. Minerals standards for electron microprobe analysis of oxygen. *Am. Mineral.* **77**, 1087–1091 (1992).
30. Gudmundsson, G. & Holloway, J. R. Activity–composition relationships in the system Fe–Pt at 1300 and 1400 °C and at 1 atm and 20 kbar. *Am. Mineral.* **78**, 178–186 (1993).
31. Van Aken, P. A., Liebscher, B. & Styrsa, V. S. Quantitative determination of iron oxidation states in minerals using Fe $L_{2,3}$ -edge electron energy-loss near-edge structure spectroscopy. *Phys. Chem. Miner.* **5**, 323–327 (1998).

Acknowledgements. We thank P. Rez and J. Bradley for discussions, and K. Righter, D. Canil and G. Gudmundsson for providing us with samples. This work was supported by the Earth Sciences Division of the US NSF.

Correspondence and requests for materials should be addressed to L.A.J.G. (e-mail: lgarvie@asu.edu).

Mice lacking melanin-concentrating hormone are hypophagic and lean

Masako Shimada*, Nicholas A. Tritos†, Bradford B. Lowell*, Jeffrey S. Flier* & Eleftheria Maratos-Flier‡

* Division of Endocrinology, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, Massachusetts 02115, USA

† Research Division, Joslin Diabetes Center, 1 Joslin Place, Boston, Massachusetts 02215, USA

Feeding is influenced by hypothalamic neuropeptides that promote (orexigenic peptides) or inhibit feeding¹. Of these, neuropeptide Y (NPY) in the arcuate nucleus² and melanin-concentrating hormone (MCH)³ and orexins/hypocretins^{4,5} in the lateral hypothalamus have received attention because their expression is increased during fasting and because they promote feeding when administered centrally. Surprisingly, absence of the orexigenic neuropeptide NPY fails to alter feeding or body weight in normal mice⁶. As deficiency of a single component of the pathway that limits food intake (such as leptin or receptors for melanocortin-4)^{7,8} causes obesity, it has been suggested that orexigenic signals are more redundant than those limiting food intake^{7,8}. To define further the physiological role of MCH and to test the redundancy of orexigenic signals, we generated mice carrying a targeted deletion of the MCH gene. MCH-deficient mice have reduced body weight and leanness due to hypophagia (reduced feeding) and an inappropriately increased metabolic rate, despite their reduced amounts of both leptin and arcuate nucleus pro-opiomelanocortin messenger RNA. Our results show that MCH is a critical

regulator of feeding and energy balance which acts downstream of leptin and the melanocortin system, and that deletion of a gene encoding a single orexigenic peptide can result in leanness.

The MCH gene was disrupted in embryonic stem cells by homologous recombination using a targeting vector in which exons 1–3 of the MCH gene were replaced by a construct consisting of phosphoglycerate kinase (PGK) promoter and the neomycin-resistance (*neo*) gene (PGK-*neo*) (Fig. 1a). Two embryonic stem cell lines with correctly targeted recombination were detected by Southern blot analysis, and a line of mice with a disrupted MCH gene (*MCH*^{−/−}) was established (Fig. 1b). Reverse transcription with polymerase chain reactions (RT-PCR; Fig. 1c) and *in situ* hybridization (Fig. 1d) confirmed disruption of the MCH gene in *MCH*^{−/−} mice. In addition, immunocytochemistry showed that MCH was not immunostained in the brain of *MCH*^{−/−} mice (data not shown).

MCH^{−/−} mice were born at the expected mendelian frequency, viable into adulthood, and fertile, and they appeared phenotypically

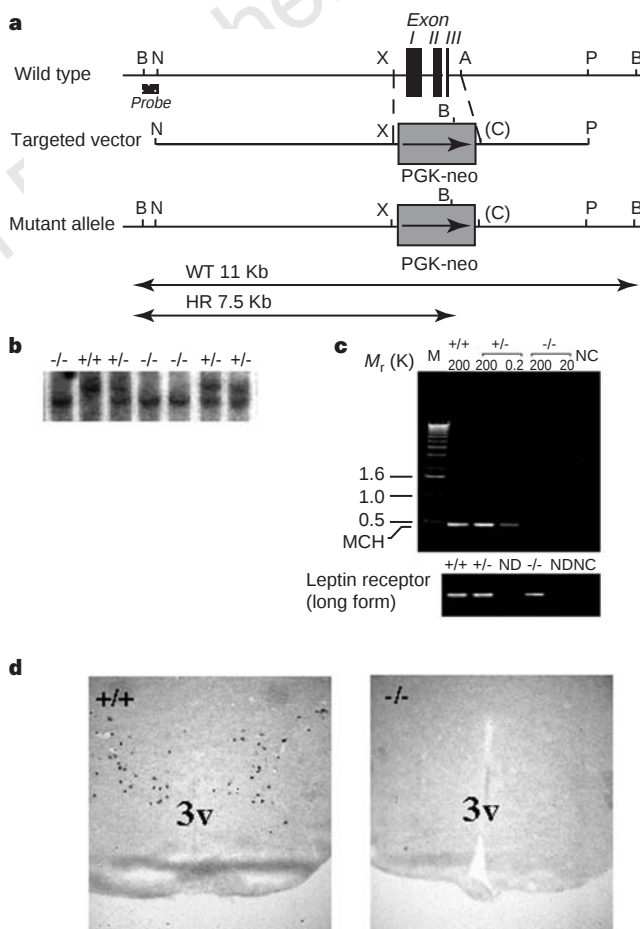


Figure 1 Targeted disruption of the MCH gene in mice. **a**, Top, the targeted region of the MCH gene locus. Middle, the PGK-*neo* cassette was used to disrupt the MCH gene coding region. Bottom, the expected targeted allele. Restriction enzyme sites: B, *Bgl*II; N, *Nco*I; X, *Xba*I; A, *Afl*I; P, *Pst*I; (C), *Cla*I in polylinker. The thick black bar indicates the genomic 320-bp *Bgl*II-*Nco*I probe. WT, wild-type; HR, the homologous recombinant allele. **b**, Southern blot analysis of mouse tail genomic DNA. Genotypes are shown at the top. **c**, Top, detection of MCH mRNA in the hypothalamus using RT-PCR. The mouse genotype and total amount of RNA (in ng) used for RT-PCR in each group is shown at the top. Bottom, RT-PCR analysis, using primers for the long form of the leptin receptor and the same RNA as in the top panel. M, marker; NC, negative control; ND, not done. **d**, *In situ* hybridization with a digoxigenin-labelled riboprobe was used to analyse MCH expression in the hypothalamus of wild-type (left) and *MCH*^{−/−} (right) mice. 3V, third ventricle.

normal by gross inspection except for their reduced size. The mean body weight of both male and female $MCH^{-/-}$ mice was reduced, and this was first apparent between 4 and 5 weeks of age. At 17 weeks of age, male $MCH^{-/-}$ mice weighed 28% less (Fig. 1a) and female $MCH^{-/-}$ mice weighed 24% less (Fig. 2b) than control mice; weights of heterozygous $MCH^{+/-}$ mice tended to be slightly but not significantly lower than weights of control mice (Fig. 2a, b). Carcass analysis showed that $MCH^{-/-}$ mice had a reduced triglyceride content; they were leaner than control littermates (Fig. 2c). We measured nose–anus length of male mice at 16 weeks of age; this measurement did not vary between groups ($MCH^{+/+}$, 9.2 ± 0.1 ; $MCH^{-/-}$, 9.1 ± 0.1 cm; $n = 6$ in each group).

The reduced body weight and lean phenotype were associated with hypophagia, as 6-week-old $MCH^{-/-}$ mice ingested 12% fewer calories than control mice over a 24-hour period (Fig. 2d). The entire reduction in food intake occurred during the dark cycle, and the amount of food consumed during the light period, although a small fraction of total daily intake, was significantly increased compared with control animals, indicating that MCH may exert its primary effect on feeding during the dark period (Fig. 2d). However, as $MCH^{-/-}$ mice consumed most of their calories during the dark cycle, like normal mice, it seems that MCH is not necessary for the normal diurnal feeding pattern (Fig. 2d). Consistent with their lean phenotype, the $MCH^{-/-}$ mice showed decreased leptin levels, being 51%, 59% and 42% of control levels at 12:00, 18:00 and

24:00, respectively ($P < 0.05$ versus control) (Fig. 3a).

As MCH has been proposed to play a role in motivated behaviours, we studied the behaviour of $MCH^{-/-}$ mice using an open-field locomotion test⁹, in which $MCH^{-/-}$ mice were as active as wild-type controls. Control male mice ($n = 6$) crossed 41 ± 6 squares over 5 minutes whereas $MCH^{-/-}$ male mice ($n = 6$) crossed 43 ± 6 squares over the same interval.

We studied potential effects of MCH deficiency on neuroendocrine function. Thyroxine levels ($MCH^{+/+}$ mice, $4.7 \pm 0.6 \mu\text{g dl}^{-1}$; $MCH^{-/-}$ mice, $5.2 \pm 1.2 \mu\text{g dl}^{-1}$; mean $\pm$ s.e.m.) and corticosterone levels ($MCH^{+/+}$ mice, $16.9 \pm 1.5 \text{ ng ml}^{-1}$; $MCH^{-/-}$ mice, $14.7 \pm 1.3 \text{ ng ml}^{-1}$) measured at 06:00 were not significantly different in 15-week-old male $MCH^{-/-}$ mice and controls. Levels of blood glucose and insulin measured in 15-week-old males were also not significantly different (glucose: $MCH^{+/+}$ mice, $107.7 \pm 5.3 \text{ mg dl}^{-1}$; $MCH^{-/-}$ mice, $102.6 \pm 5.9 \text{ mg dl}^{-1}$; insulin: $MCH^{+/+}$ mice, $0.99 \pm 0.06 \text{ ng ml}^{-1}$; $MCH^{-/-}$ mice, $0.72 \pm 0.07 \text{ ng ml}^{-1}$; $n = 7$ each group).

The reduction in body weight in $MCH^{-/-}$ mice exceeded the degree to which food intake was reduced, indicating that increased metabolic rate might contribute to the lower weight. Rectal temperature, an imperfect indicator of metabolic rate, was unchanged, being $37.9 \pm 0.03^\circ\text{C}$ in control animals ($n = 6$) and $37.6 \pm 0.04^\circ\text{C}$ in $MCH^{-/-}$ animals ($n = 6$). We therefore measured the metabolic rate of these mice in the fed state. $MCH^{-/-}$ mice show rates of oxygen consumption that are slightly but not significantly higher than control rates when expressed on a per mouse basis ($MCH^{+/+}$ mice, $1.08 \pm 0.06 \text{ ml O}_2$ per min per mouse; $MCH^{-/-}$ mice, $1.14 \pm 0.06 \text{ ml O}_2$ per min per mouse). However, the rate of oxygen consumption is increased by 20% in $MCH^{-/-}$ mice when normalized to body mass ($MCH^{-/-}$ mice, $34.1 \pm 0.4 \text{ ml O}_2$ per min per kg body weight; $MCH^{+/+}$ mice, $27.8 \pm 0.4 \text{ ml O}_2$ per min per kg body weight). Whether expressed per mouse or per kg body weight, this rate of

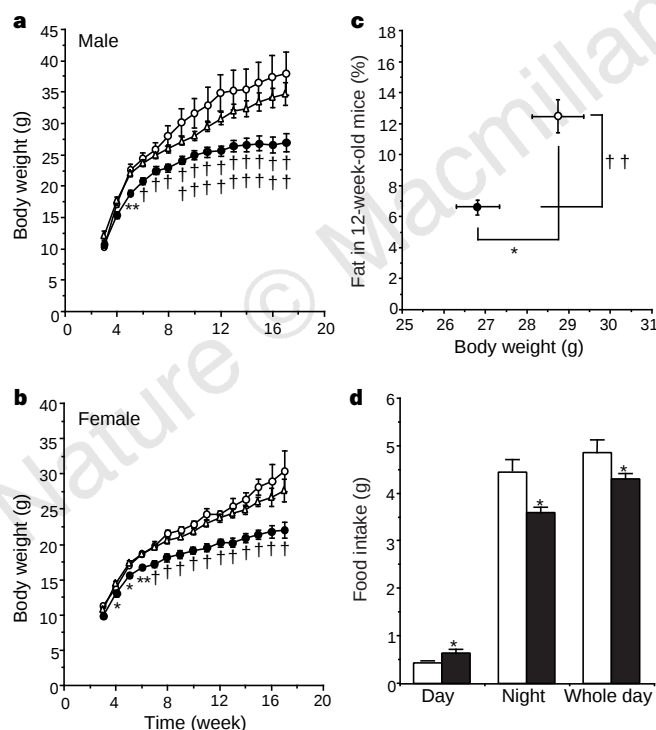


Figure 2 Body weight, fat composition and food intake in $MCH^{-/-}$ and control mice fed on a chow diet. **a, b**, Growth curves. Three to four mice were housed per cage. Male: $MCH^{+/+}$, $n = 8$; $MCH^{+/-}$, $n = 19$; $MCH^{-/-}$, $n = 13$. Female: $MCH^{+/+}$, $n = 6$; $MCH^{+/-}$, $n = 28$; $MCH^{-/-}$, $n = 12$. **c**, Percentage of triglycerides relative to body weight in $MCH^{-/-}$ and $MCH^{+/+}$ mice. Open circles, wild-type mice; open triangles, $MCH^{+/-}$ mice; shaded circles, $MCH^{-/-}$ mice. **d**, Food intake by 6-week-old wild-type ($n = 6$; white bars) and $MCH^{-/-}$ ($n = 9$; shaded bars) male mice during the light and dark cycles. Mice were housed individually 4 days before the experiment and the values were determined over a 5-day period. Data are mean intake per day $\pm$ s.e.m. Asterisk indicates $P < 0.05$; double asterisk indicates $P < 0.01$; dagger indicates $P < 0.001$; double dagger indicates $P < 0.0001$. Statistical significance was determined by Fischer's PLSD (post-hoc least significant difference).

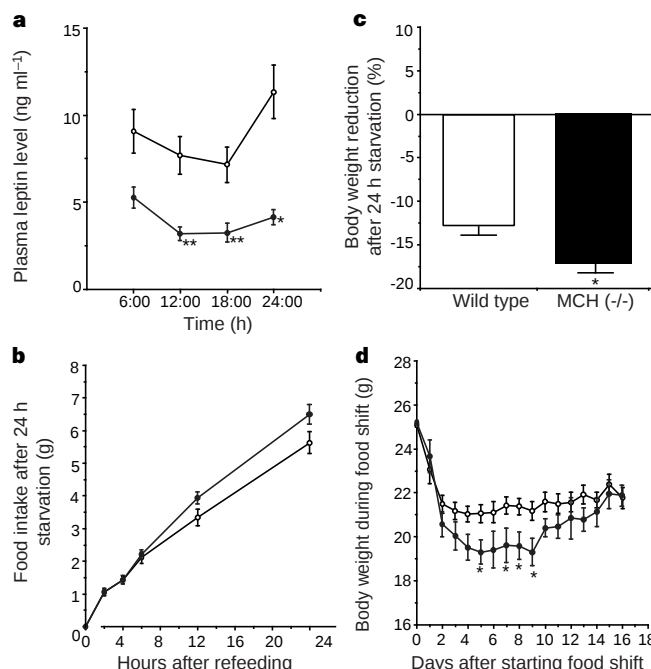


Figure 3 Leptin levels and feeding response in $MCH^{-/-}$ mice. **a**, Leptin levels in 8-week-old male $MCH^{-/-}$ ($n = 6$; filled circles) and $MCH^{+/+}$ ($n = 8$; open circles) mice. Data are means $\pm$ s.e.m. **b**, Cumulative food intake in response to starvation in $MCH^{-/-}$ (filled circles) and $MCH^{+/+}$ (open circles) mice. **c**, Body-weight reduction after 24-h starvation. **d**, Response of wild-type (open circles) and $MCH^{-/-}$ (filled circles) mice to food shift. Asterisk indicates $P < 0.05$; double asterisk indicates $P < 0.01$. Statistical significance was determined by Fischer's PLSD.

oxygen consumption may reflect an impaired ability of $MCH^{-/-}$ mice to appropriately regulate their metabolic rate in response to their reduced feeding and lower amounts of leptin (see below). Thus, MCH deficiency promotes weight loss by reducing feeding and may also limit the normal suppression of metabolic rate during food restriction. This combination of results was predicted for $NPY^{-/-}$ mice on the basis of the results of intracerebral injections of this neuropeptide, but mice deficient in NPY showed no such defects⁶.

We studied the response of $MCH^{-/-}$ mice to food deprivation. $MCH^{-/-}$ mice responded to 24 hours of starvation with compensatory hyperphagia equal to that of wild-type mice (Fig. 3b). Thus, like NPY^0 , MCH is not essential for the hyperphagic response to starvation. However, two lines of evidence indicate that MCH deficiency may create susceptibility to the adverse effects of starvation. First, when age-matched mice were starved for 24 hours, weight loss was greater in $MCH^{-/-}$ mice than in controls ($17.0 \pm 1.0\%$ vs $12.8 \pm 0.9\%$ weight loss, respectively; $n = 6$, $P < 0.05$) (Fig. 3c). Second, when age-matched mice were starved for 48 hours, 3 of 4 $MCH^{-/-}$ but 0 of 4 $MCH^{+/+}$ mice died during the last 4 hours of the fast. The precise basis for these adverse effects of starvation is not yet known, but inappropriate levels of thermogenesis, coupled with reduced initial fat stores, would be expected to cause greater weight loss, and earlier death, in the absence of food intake.

To further explore the response to altered food availability, we used a restricted-feeding paradigm¹⁰. Food was offered only during the light cycle, between 11:00 and 15:00, to 7-week-old mice. Over a 9-day period, wild-type controls responded to this stress with a 16% reduction in body weight, whereas $MCH^{-/-}$ mice showed 23% weight loss ($P < 0.05$) (Fig. 3d). In contrast, food intake fell to the same degree in wild-type and $MCH^{-/-}$ mice (data not shown), consistent with the possibility that MCH deficiency causes increased thermogenesis under these circumstances.

The hypophagia and leanness of $MCH^{-/-}$ mice led to the production of lower leptin levels, which should cause hyperphagia. However, in the context of MCH deficiency, low leptin levels resulted in decreased rather than increased food intake, indicating that MCH may be essential for the hyperphagic response to leptin deficiency under these conditions. We also examined the effect of exogenous

leptin on $MCH^{-/-}$ mice. Leptin was injected at a dose of $1 \mu\text{g}$ per g body weight twice a day for 2 days. Leptin was able to reduce both food intake and body weight¹¹ within the initial 24-hour period (Fig. 4). The initial response of $MCH^{-/-}$ mice to leptin was twice that of wild-type mice ($P < 0.05$). This exaggerated response to exogenous leptin, which was also seen in $NPY^{-/-}$ mice, was attenuated by day 2. As leptin inhibits food intake in $MCH^{-/-}$ mice, suppression of MCH is unlikely to be the mechanism by which leptin inhibits food intake in normal mice. Furthermore, MCH may, either directly or indirectly, oppose the effect of leptin in diminishing food intake.

Finally, to assess the state of hypothalamic compensation in response to both MCH deficiency and lower leptin levels, we studied hypothalamic expression of mRNAs encoding four other appetite-regulating neuropeptides in $MCH^{-/-}$ mice, namely NPY, orexin⁴, Agouti-related protein (AGRP)^{12,13} and pro-opiomelanocortin (POMC)^{14,15}. Expression of NPY and AGRP in the arcuate nucleus and of orexin in the lateral hypothalamus remained unchanged in $MCH^{-/-}$ mice in the fed state. However, expression of POMC in the arcuate nucleus was markedly suppressed (decreased by 63% compared with controls; $P < 0.006$) (Fig. 5). As leptin is a positive regulator of arcuate POMC expression^{16,17} the reduced POMC expression could result from lower leptin levels, which, in turn, result from a lean body composition. Lower leptin amounts did not increase NPY expression in $MCH^{-/-}$ mice but did in wild-type mice¹⁸. These results might indicate that POMC expression is more sensitive to the reduction in leptin levels than is NPY expression. Alternatively, MCH deficiency could reduce arcuate POMC expression directly, through the loss of direct innervation of arcuate POMC-expressing neurons by MCH-expressing neurons; MCH immunoreactivity is seen in the arcuate nucleus of wild-type mice¹⁹, consistent with bidirectional communication between POMC- and MCH-expressing systems. There are bidirectional connections between arcuate neurons expressing NPY, AGRP and POMC and lateral hypothalamic neurons expressing MCH and orexins²⁰. The lower expression of arcuate POMC, which in other contexts promotes hyperphagia and obesity²¹, might be expected to limit the consequences of MCH deficiency. However, our results show that the ability of low arcuate POMC expression (and/or a low POMC:AGRP ratio) to cause hyperphagia and obesity is limited in the absence of an intact MCH gene.

Our $MCH^{-/-}$ mice have a deletion of the entire coding region of the MCH gene, which also encodes the neuropeptides EI and GE²². There is little information available on the biological role of these two peptides; direct administration of either peptide alone or in combination in rats has no effect on eating behaviour (E.M.-F., unpublished observations). In theory, the phenotypes of $MCH^{-/-}$ mice could result from deficiency of any or all of these peptides.

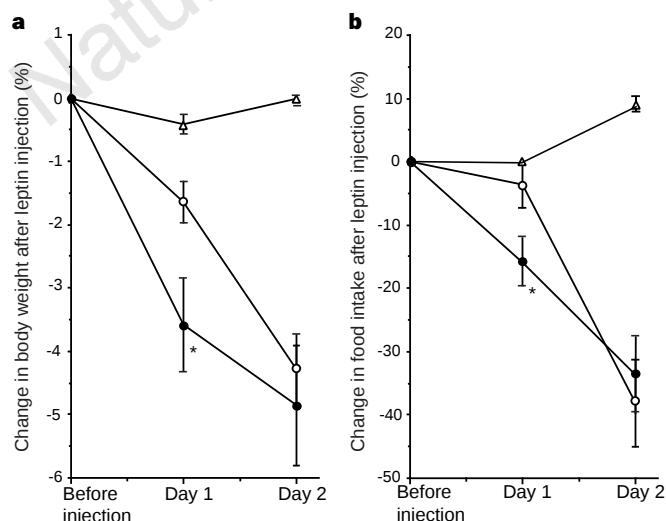


Figure 4 Response of six $MCH^{+/+}$ (open circles) and six $MCH^{-/-}$ mice (shaded circles) to recombinant mouse leptin, administered at a dose of $1 \mu\text{g}$ per g body weight twice daily (09:00 and 18:00), vs saline control ($n = 3$; triangles). Values for **a**, the change in body weight and **b**, the change in food intake are the per cent change from a 7-day baseline period. Asterisk indicates $P < 0.05$. Statistical significance was determined by Fischer's PLSD.

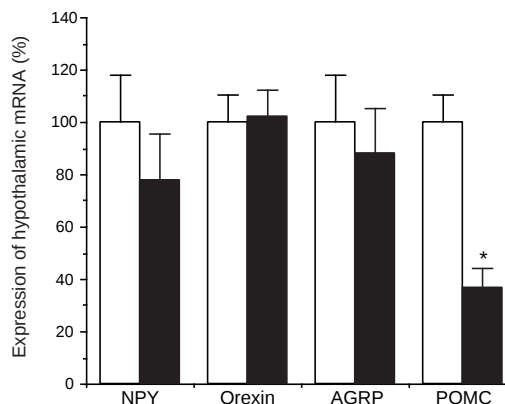


Figure 5 Expression of neuropeptide mRNAs in hypothalami of 24-week-old wild-type ($n = 6$; white bars) and $MCH^{-/-}$ ($n = 6$; shaded bars) mice. Asterisk indicates $P < 0.006$. Statistical significance was determined with the Mann-Whitney test.

However, given the ability of MCH to promote feeding^{3,23} and the absence of evidence that NEI and NGE promote such an effect, it is reasonable to assume that MCH deficiency is the primary, or even sole, cause of the phenotype reported here.

MCH is a cyclic 19-amino-acid polypeptide whose expression is limited to the lateral hypothalamus and zona incerta^{22,24}. It has been implicated as an important regulator of eating behaviour, because central administration of MCH promotes feeding, and MCH mRNA amounts rise as a result of starvation and leptin deficiency³. MCH-expressing neurons are well placed anatomically to participate in feeding behaviour, because they make monosynaptic connections with several areas in the brain involved in integrating inputs related to taste, olfaction and visceral sensations, including the nucleus of the solitary tract, the parabrachial nucleus and the insular and medial prefrontal cortex²⁴. MCH-expressing neurons have therefore been suspected to participate in complex integrative behaviours, and the acute feeding response to MCH^{3,23} indicated that feeding was among these behaviours. The phenotype of the *MCH*^{-/-} mouse reported here strongly supports this hypothesis. As the leanness of *MCH*^{-/-} mice occurs despite reduced expression of leptin and arcuate POMC, which produce obesity when MCH is present^{7,21,25}, MCH appears to be a critical effector of energy balance downstream of leptin and the POMC/melanocortin system. Further genetic crosses between *MCH*^{-/-} mice and mice with different causes of obesity will be required to test this hypothesis directly. Finally, as *MCH*^{-/-} mice are lean, antagonists of MCH action may be effective treatments for obesity.

Methods

Creation of *MCH*^{-/-} mice. We screened a mouse C129 SvJ P1 genomic library (Genome Systems) using a 250-base-pair (bp) PCR fragment, generated on the basis of the published sequence of the murine MCH gene²⁶, as a probe. One of three clones was mapped; it consisted of the 16-kilobase (kb) P1 vector and a 72-kb genomic insert containing the MCH coding region between ~25 kb of 5' and 45 kb of 3' flanking sequence. We replaced a ~1.8 kb-fragment extending from 340 bp upstream of the translation initiation codon (*Xba*I site) and 200 bp downstream of the stop codon (*Afl*III site) of this clone with a *PGK-neo*^r cassette. This targeting vector in pGEM5 (Promega) was linearized at the *Sall* site and electroporated into the J1 embryonic stem line²⁷ (cells provided by E. Li and A. Sharp); selection of G418-resistant clones was done as described²⁷. We identified targeted clones by Southern blot analysis, using a 320-bp *Bgl*II-*Nco*I fragment located 5' of the targeting vector as a probe; 2 of 68 clones were positive. These positive clones were injected into C57B1/6 embryos at the blastocyst stage. Chimaeric offspring were mated with C57B1/6 mice. Germline transmission of the mutant allele was determined by Southern analysis of mouse tail genomic DNA. One line of mice carrying the disrupted MCH gene was generated, and F₃ hybrids were used in all experiments. We reverse-transcribed 1 µg of total RNA from the hypothalamus of a mouse from each group and amplified aliquots equivalent to 200, 20 and 0.2 ng total RNA by PCR, using the following primers: sense, 5'-ATGGCAAGATGACTCTCTCT-3'; and antisense, 5'-GACTTGCCAACATGGTCGGTA-3', as described²⁶. *In situ* hybridization was done using a digoxigenin-labelled MCH complementary RNA probe²⁸.

Studies of *MCH*^{-/-} mice. Mice were maintained in a daily cycle of 12 h light (06:00–18:00) and 12 h darkness (18:00–06:00) and were allowed free access to chow (Purina Formulab 5008) and water. Total body lipid content was assessed using alcoholic potassium hydroxide digestion with saponification of all fats, neutralization and then enzymatic determination of glycerol (Sigma) as described²⁹. Leptin and insulin concentrations were measured by RIA (Linco Research), as were corticosterone and thyroxine levels (ICN). Glucose was measured using a One Touch meter (Lifescan). Rectal temperature was measured with a rectal probe (Yellow Springs Instrument). Oxygen consumption was measured in 20-week-old male wild-type (*n* = 4) and *MCH* (*n* = 4) mice using an Oxymax 4.93 (Columbus Instruments).

Response to starvation in *MCH*^{-/-} and *MCH*^{+/-} mice. We deprived 10-week-old mice (*n* = 6) of food for 24 h, beginning at 09:00; the mice were then allowed free access to chow. To measure cumulative food intake after refeeding,

we measured the weight of consumed food at 2, 4, 6, 12 and 24 h after refeeding. Body weights were measured after the 24-h starvation period and after refeeding.

Food-shift experiment. To determine the response of wild-type and *MCH*^{-/-} mice to food shift, body-weight-matched animals (wild-type male, *n* = 7; *MCH*^{-/-} male, *n* = 4) were individually housed and allowed free access to chow for only 4 h during the light cycle (11:00–15:00) for 16 days¹⁰. Body weights were measured at 11:00. Animals were allowed free access to water during the period.

Response to leptin. To determine the response to leptin, mice were injected intraperitoneally with recombinant mouse leptin (Eli Lilly) at a dose of 1 µg per g body weight twice daily (09:00 and 18:00). Three control mice were injected with saline.

Expression of hypothalamic mRNAs. Mice were killed at 09:00 by administration of pentobarbital and were perfused transcardially first with saline and then with 10% buffered formalin. *In situ* hybridization was done with probes for NPY, POMC, AGRP and orexin mRNAs. The images captured on film were quantified with a computing densitometer (Molecular Dynamics) and ImageQuant Software (Molecular Dynamics). We integrated the absorbance in rectangular areas encompassing each nucleus of interest (arcuate nucleus for NPY, POMC and AGRP, and lateral hypothalamus for orexin) over each set of brain sections, and subtracted background density from an adjacent area without specific hybridization signal. We determined statistical significance with the Mann–Whitney test.

Received 15 July; accepted 5 October 1998.

1. Flier, J. S. & Maratos-Flier, E. Obesity and the hypothalamus: novel peptides for new pathways. *Cell* **92**, 437–440 (1998).
2. Dryden, S., Frankish, H., Wang, Q. & Williams, G. Neuropeptide Y and energy balance: one way ahead for the treatment of obesity? *Eur. J. Clin. Invest.* **24**, 293–308 (1994).
3. Qu, D. *et al.* A role for melanin-concentrating hormone in the central regulation of feeding behaviour. *Nature* **380**, 243–247 (1996).
4. Sakurai, T. *et al.* Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell* **92**, 573–585 (1998).
5. de Lecea, L. *et al.* The hypocretins: hypothalamus-specific peptides with neuroexcitatory activity. *Proc. Natl Acad. Sci. USA* **95**, 322–327 (1998).
6. Erickson, J., Clegg, K. & Palmiter, R. Sensitivity to leptin and susceptibility to seizures of mice lacking neuropeptide Y. *Nature* **381**, 415–418 (1996).
7. Zhang, Y. *et al.* Positional cloning of the mouse ob gene and its human homologue. *Nature* **372**, 425–432 (1994).
8. Huszar, D. *et al.* Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88**, 131–141 (1997).
9. Krezel, W. *et al.* Impaired locomotion and dopamine signaling in retinoid receptor mutant mice. *Science* **279**, 863–867 (1998).
10. Ahima, R. S., Prabakaran, D. & Flier, J. S. Postnatal leptin surge and regulation of circadian rhythm of leptin by feeding. Implications for energy homeostasis and neuroendocrine function. *J. Clin. Invest.* **101**, 1020–1027 (1998).
11. Halaas, J. *et al.* Weight reducing effect of the plasma protein encoded by the obese gene. *Science* **269**, 543–546 (1995).
12. Ollmann, M. M. *et al.* Antagonism of central melanocortin receptors *in vitro* and *in vivo* by Agouti-related protein. *Science* **278**, 135–138 (1997).
13. Shutter, J. R. *et al.* Hypothalamic expression of ART, a novel gene related to agouti, is up-regulated in obese and diabetic mutant mice. *Genes Dev.* **11**, 593–602 (1997).
14. Tsujii, S. & Bray, G. A. Acetylation alters the feeding response to MSH and beta-endorphin. *Brain Res. Bull.* **23**, 165–169 (1989).
15. Fan, W., Boston, B. A., Kesterson, R. A., Hruby, V. J. & Cone, R. D. Role of melanocortinergic neurons in feeding and the agouti obesity syndrome. *Nature* **385**, 165–168 (1997).
16. Thornton, J. E., Cheung, C. C., Clifton, D. K. & Steiner, R. A. Regulation of hypothalamic proopiomelanocortin mRNA by leptin in ob/ob mice. *Endocrinology* **138**, 5063–5066 (1997).
17. Mizuno, T. M. *et al.* Hypothalamic pro-opiomelanocortin mRNA is reduced by fasting in ob/ob and db/db mice, but is stimulated by leptin. *Diabetes* **47**, 294–297 (1998).
18. Ahima, R. S. *et al.* Role of leptin in the neuroendocrine response to fasting. *Nature* **382**, 250–252 (1996).
19. Bittencourt, J. C. & Elias, C. F. Diencephalic origins of melanin-concentrating hormone immunoreactive projections to medial septum/diagonal band complex and spinal cord using two retrograde fluorescent tracers. *Ann. NY Acad. Sci.* **680**, 462–465 (1993).
20. Elias, C. F. *et al.* Chemically defined projections linking the mediobasal hypothalamus and the lateral hypothalamic area. *J. Comp. Neurol.* (in press).
21. Krude, H. *et al.* Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nature Genet.* **19**, 155–157 (1998).
22. Nahon, J. L. The melanin-concentrating hormone: from the peptide to the gene. *Crit. Rev. Neurobiol.* **8**, 221–262 (1994).
23. Rossi, M. *et al.* Melanin-concentrating hormone acutely stimulates feeding, but chronic administration has no effect on body weight. *Endocrinology* **138**, 351–355 (1997).
24. Bittencourt, J. C. *et al.* The melanin-concentrating hormone system of the rat brain: an immunohistochemical characterization. *J. Comp. Neurol.* **319**, 218–245 (1992).
25. Montague, C. T. *et al.* Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature* **387**, 903–908 (1997).
26. Breton, C., Presse, F., Hervieu, G. & Nahon, J. L. Structure and regulation of the mouse melanin-concentrating hormone mRNA and gene. *Mol. Cell. Neurosci.* **4**, 271–284 (1993).
27. Li, E., Sucov, H. M., Lee, K. F., Evans, R. M. & Jaenisch, R. Normal development and growth of mice carrying a targeted disruption of the alpha 1 retinoic acid receptor gene. *Proc. Natl Acad. Sci. USA* **90**, 1590–1594 (1993).
28. Marks, D. L. *et al.* Simultaneous visualization of two cellular mRNA species in individual neurons by use of a new double *in situ* hybridization method. *Mol. Cell. Neurosci.* **3**, 395–405 (1993).

29. Lowell, B. B. *et al.* Development of obesity in transgenic mice after genetic ablation of brown adipose tissue. *Nature* **366**, 740–742 (1993).

Acknowledgements. We thank J. Mastaitis, C. Behn and C. Lee for technical assistance, J. Elmquist for help in analysing and interpreting brain anatomy, and D. S. Ludwig for the P1 clone used in making the construct for the knockout mice. This work was supported in part by grants from NIH to J.S.F. and E.M.-F., from the American Diabetes Association to E.M.-F., and from Eli Lilly to J.S.F. and E.M.-F., and by the Transgenic Core of the Boston Obesity Nutrition Research Center. M.S. was supported by the Banyu Fellowship in Lipid Metabolism and Atherosclerosis which is sponsored by Banyu Pharmaceutical Co Ltd and the Merck Foundation.

Correspondence and requests for materials should be addressed to E.M.-F. (e-mail: emarat@joslab.harvard.edu).

GABA_B receptors function as a heteromeric assembly of the subunits GABA_BR1 and GABA_BR2

Kenneth A. Jones, Beth Borowsky, Joe A. Tamm, Douglas A. Craig, Margaret M. Durkin, Meng Dai, Wen-Jeng Yao, Mary Johnson, Caryn Gunwaldsen, Ling-Yan Huang, Cheng Tang, Quanrong Shen, John A. Salon, Kelley Morse, Thomas Laz, Kelli E. Smith, Dhanapalan Nagarathnam, Stewart A. Noble, Theresa A. Branchek & Christophe Gerald

Synaptic Pharmaceutical Corporation, 215 College Road, Paramus, New Jersey 07652, USA

The principal inhibitory neurotransmitter GABA (γ -aminobutyric acid) exerts its effects through two ligand-gated channels, GABA_A and GABA_C receptors, and a third receptor, GABA_B (ref. 1), which acts through G proteins to regulate potassium and calcium channels. Cells heterologously expressing the cloned DNA encoding the GABA_BR1 protein exhibit high-affinity antagonist-binding sites², but they produce little of the functional activity expected from studies of endogenous GABA_B receptors in the brain. Here we describe a new member of the GABA_B polypeptide family, GABA_BR2, that shows sequence homology to GABA_BR1. Neither GABA_BR1 nor GABA_BR2, when expressed individually, activates GIRK-type potassium channels; however, the combination of GABA_BR1 and GABA_BR2 confers robust stimulation of channel activity. Both genes are co-expressed in individual neurons, and both proteins co-localize in transfected cells. Moreover, immunoprecipitation experiments indicate that the two polypeptides associate with each other, probably as heterodimers. Several G-protein-coupled receptors (GPCRs) exist as high-molecular-weight species, consistent with the formation of dimers by these receptors^{3–7}, but the relevance of these species for the functioning of GPCRs has not been established. We have now shown that co-expression of two GPCR structures, GABA_BR1 and GABA_BR2, belonging to the same subfamily is essential for signal transduction by GABA_B receptors.

To find other genes related to the GABA_BR1 gene, we searched the expressed sequence tags (ESTs) of GenBank using the GABA_BR1 sequence. (Throughout the text, 'GABA_BR1' refers to the GABA_BR1b splice variant.) Two entries had scores that suggested significant homology to GABA_BR1. We used oligonucleotide probes from these sequences to isolate a full-length clone from a rat hypothalamic complementary DNA library. Sequence analysis of this new clone showed that it has a 2.8-kilobase open reading frame that is predicted to encode a protein of 940 amino acids. A BLAST search of the GenEMBL database indicated that this amino-acid sequence was most closely related to that of GABA_BR1, exhibiting 35% and 41% identity overall and within the predicted transmembrane domains, respectively (Fig. 1a). The structural similarity to

GABA_BR1 indicated that this sequence might encode a new GABA_B polypeptide, which we refer to as GABA_BR2. The next most related sequences were other members of the metabotropic glutamate receptor (mGluR) family, with 21–24% overall amino-acid identity. Like GABA_BR1 and other members of the mGluR family⁸, GABA_BR2 contains a large amino-terminal extracellular domain with regions of homology to bacterial amino-acid-binding proteins.

We studied the distribution of GABA_BR2 messenger RNA within the central nervous system by *in situ* hybridization. Strong hybridization signals were observed in regions of the rat brain that have high densities of GABA_B-receptor-binding sites⁹, such as the hippocampus, medial habenula, thalamus and cerebellum (Fig. 1b–e). There was a high degree of similarity in the distribution and intensity of GABA_BR1 and GABA_BR2 hybridization signals.

Postsynaptic inhibition of neurons by GABA_B-receptor activation is caused by the opening of inwardly rectifying K⁺ channels (GIRKs)^{10–13}. We assessed the ability of either GABA_BR1 or GABA_BR2 to regulate K⁺ currents in cells transfected with GIRK1 and GIRK4 subunits^{14–16}. *Xenopus* oocytes injected with either GABA_BR1 or GABA_BR2 mRNAs failed to generate GABA-evoked GIRK currents (Fig. 2b). The longer splice variant of GABA_BR1, GABA_BR1a, did not stimulate GIRK activity either, as reported elsewhere^{2,17}. In a mammalian host, HEK293 cells, small agonist-evoked currents (10–50 pA) were observed in 5 of 26 cells expressing GABA_BR1; similar weak currents were evoked in 1 of 23 cells expressing GABA_BR2. In contrast, large currents were produced in 100% of cells expressing the galanin receptor GalR1 (ref. 18) (Fig. 2i).

The overlapping expression patterns of GABA_BR1 and GABA_BR2 transcripts in the brain indicated that the corresponding proteins might be co-expressed in individual neurons and that both might be required for functional activity. When GABA_BR1 and GABA_BR2 were co-injected into oocytes together with GIRK subunits, the application of GABA produced robust K⁺ currents (Fig. 2a, b). Responses to GABA were abolished in oocytes pretreated with pertussis toxin, showing that receptor stimulation leads to activation of a heterotrimeric G protein of the G α_o /G α_i class. The GABA-induced currents showed the following properties, which suggest that the currents were mediated by GIRK channels: first, a dependency on a raised external K⁺ concentration; second, strong inward rectification; third, a reversal potential (–23 mV) near the predicted equilibrium potential for K⁺; and fourth, sensitivity to block by 100 μ M Ba²⁺ (Fig. 2c).

The pharmacology of agonists at the GABA_BR1/GABA_BR2 combination was comparable to that reported for native receptors^{19,20}. GABA, baclofen and 3-aminopropylmethyl phosphinic acid (3-APMPA) exhibited half-maximally effective concentrations (EC₅₀ values) of 1.3 ± 0.1 ($n = 20$), 3.3 ± 0.4 ($n = 8$), and 0.051 ± 0.003 ($n = 6$) μ M, respectively (Fig. 2d, g). Concentration–effect curves for GABA were shifted to the right, in an apparently competitive manner, by well-characterized GABA_B-selective antagonists (Fig. 2h). Estimates of affinity of the antagonists CGP54626 (33.1 ± 6.3 nM; $n = 4$) and CGP55845 (2.5 ± 0.2 nM; $n = 6$) for GABA_BR1/GABA_BR2 were similar to values reported in previous electrophysiological studies using brain tissue^{19,20}, as well as to those obtained by measuring displacement of radioligand from cells expressing GABA_BR1 alone².

The appearance of robust functional responses after GABA_BR1/GABA_BR2 co-expression was also observed in HEK293 cells. Upon co-expression of GABA_BR1 and GABA_BR2 together with GIRKs, GABA evoked currents in 70 of 81 recorded cells (Fig. 2e, i). These currents were blocked by low concentrations of the competitive GABA_B-receptor antagonist CGP55845 (ref. 21) and by the GIRK-channel blocker Ba²⁺ (Fig. 2f), indicating that they resulted from the stimulation of GABA_B receptors and the subsequent activation of GIRKs. Large-amplitude currents were also observed when GABA_BR2 was paired with the GABA_BR1a splice variant ($1,046 \pm 247$ pA; $n = 9$). To

a

GABA_BR2 MASPPSSGQPRPPPPPPARLLPLLLSLLWLAPGAWGTRGAPRPPSSPP...LSIMGLMPLTK 65
 GABA_BR1bMGPGGPCTPVGWPLPLLLVMAAGVAPVWASHSPHLPRPHRPVPPHPSSERRAVYIGALFP 60

GABA_BR2 EVAKGSIGRGVLPVELAIEQIRN.ESLLRPYFLDLRLYDTECDNAKGLKAFYDAIKYGNHLMVFGGVC 134
 GABA_BR1b MSGGWPGGQACQPAVEMALEDVNSRRDILPDYELKLIHDSKCDPGQATKYLYELLYNDPIKILMPG.C 129

GABA_BR2 PSVTISIAESLQGNLVQLSFAATTPLVADKKKYPPFFRTVPSDNAVNPAILKLLKHFWRVRVGLTQDV 204
 GABA_BR1b SSVSTLVAEARMWNLIVLSYGSSPALSNRQRFPTFFRTHPSATLHNPTRVKLFEKGWKKIATIQQT 199

GABA_BR2 QRFSEVRNDLTGVLYGEDIETSDTESFNDPCTSVKKLKGNVDRIILGQFDQNMMAKVCFCAFEESMFGS 274
 GABA_BR1b EVFTSTLDDLEERVKEAGIEITFRQSFSDPAVPVKNLKRQDARIIVGLFYETEARKVFCVEYKERLFGK 269

GABA_BR2 KYQWIIPGWYEPAWWEQVHVEANSSRLRRSLAAMEGYIGVDFEPLSSKQIKTISGKTPQQYEREYNSK 344
 GABA_BR1b KYVWFLIGWYADNWFKYDPSIN...CTVEEMTEAVEGHITTEIVMLNPANTRISNMTSQEFV.EKLTK 335

GABA_BR2 RSGVGPSKFHY...AYDGIWIAKTQRAMETLHASSRHQRIQDFNYTDHTLGKILNAMNETNFFG 409
 GABA_BR1b RLKRHPEETGGFQEAPLAYDAIWALALANKTSGGGGSG...VRLEDFNYNQITITDQTYRAMNSSSEF 403

GABA_BR2 VTGQVVF.RNGERMGTIKFTQFQDSREVKGVEYNAVADTLEIINDTIRFQGSPPKDKTIIILEQLRKISL 478
 GABA_BR1b VSGHVVDASGSRMAWTLIEQLQGGSYKKIGYYDSTKDDL.SWKTDKWIGGSPPADQTLVIKTFRFLSQ 472

GABA_BR2 PLYSILSALTILGMIMASAFFFNKRNQKLKMSPPYMNLIILGGMLSYASIFLGLDGSFVSEKTF 548
 GABA_BR1b KLFISVSVLSSLGIVLAVVCLSFNIYNHVRVYIQNSQPNLNLTAVGCSLALAFAVPLGLDGYHIGRSQF 542

GABA_BR2 ETLCVTRTWILTVGYTTAFGAMFAKTWRVHAIFKNVKKK...KIIKDQKLLVIVGGMLLIDLCLICWQ 615
 GABA_BR1b PFVCQARLWLLGLGFSGLYGSMTFKIWWVHTVFTKKEEKEWRKTLPEWKLYATVGLLVGMDVLTIAIWQ 612

GABA_BR2 AVDPLRRTVERYSMEDPDAGRDISIRPLLEHCENTHMTIWLGIYVAYKGLLMLFGCFLAWETRNVSIPAL 685
 GABA_BR1b IVDPLHRTIETFAKEPKEDIDVSLPQLEHCSSKKMNTWLGIFYGYKGLLLGLIFLAYETKSVSTEKI 682

GABA_BR2 NDSKYIGMSVYNVGIMCIIGAASFLTRDQPNVQFCIVALVIFCSTITLCLVFVPKLTITLRTNPDAATQ 755
 GABA_BR1b NDHRAVGMAIYNVAVLCITAPVTMILSSQDAAFASLAIVFSSYITLVLFVPMRRLITRGE... 748

GABA_BR2 NRRFQFTQNKKEDSKTSTSVNQASTSRLEGLQSENHRLRMKITELDKDLEEVMTQLQDTPKTTYI 825
 GABA_BR1bWQSETQDTMKTGSS.TNNNEEEKSRL..LEKNRELEKIIAEKEERVSELRHQLQRQQLRSRR 809

GABA_BR2 KQNHQELNDILSLGNFTSTDDGKAILKNHLDQNPQLQWNTTEPSRTCKDPIEDINSPEHIQRRLSLQL 895
 GABA_BR1b HPPTPPDPSGGLPRGPSEPPDRLSCDGSRVHLLYK*..... 845

GABA_BR2 PILHHAYLPSIGGVDASCVSPCVPTASPRHRHVPSPFRVMVSGL*..... 940

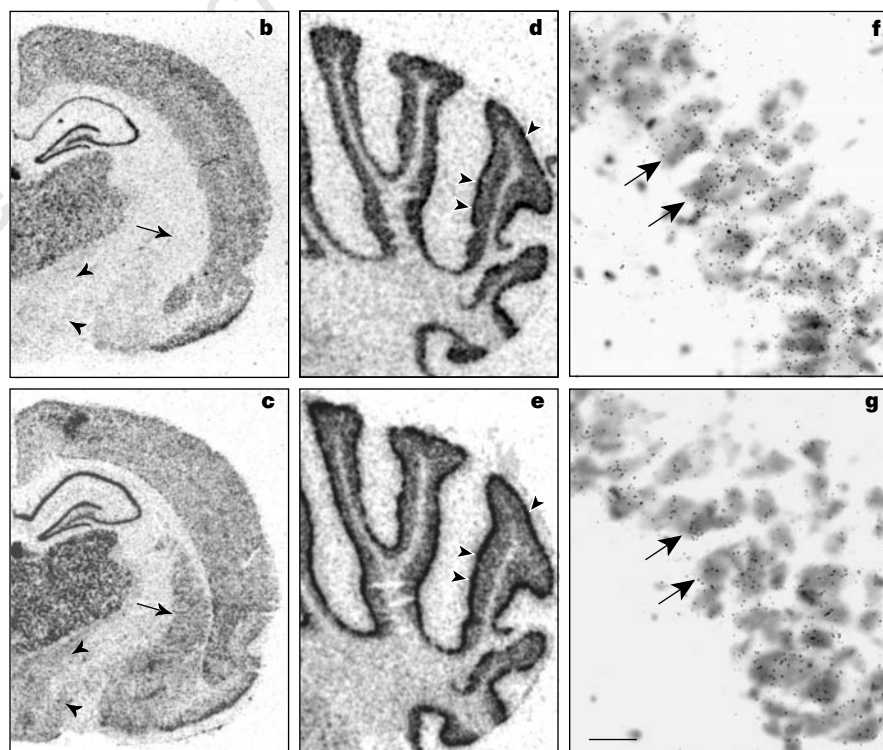


Figure 1 Amino-acid sequence of GABA_BR2 and localization by *in situ* hybridization. **a**, Alignment of the predicted GABA_BR2 and GABA_BR1b proteins. Proposed sites of signal-peptide cleavage are indicated by arrowheads. Shaded regions highlight sequence identities. Horizontal bars indicate transmembrane regions. **b-e**, Photomicrographs showing the regional distribution of GABA_BR2 (**b, d**) and GABA_BR1 (**c, e**) mRNAs in pairs of adjacent coronal rat brain sections. **b**,

c, Hypothalamus and caudate-putamen are identified with arrowheads and arrows, respectively. **d, e**, Arrowheads identify the Purkinje cell layer in the cerebellum. **f, g**, High-magnification views of hippocampal CA3 region from alternate sections, showing co-localization of GABA_BR2 transcripts (**f**) and GABA_BR1 transcripts (**g**) in the same cells (arrows). Scale bar, 30 μm.

determine whether co-expressed GABA_BR1/GABA_BR2 could mediate a cellular response independently of exogenously supplied GIRKs, we used a microphysiometer to measure agonist-evoked extracellular acidification in CHO cells expressing GABA_BR1 and GABA_BR2. Baclofen stimulated a ninefold increase in acidification rate (Fig. 2j), which was blocked by 100 nM CGP55845 and by pretreatment with pertussis toxin (data not shown). This response was absent in cells expressing either protein alone. As GIRK activity is undetectable in wild-type CHO cells¹⁵ we conclude that GIRK expression is not a prerequisite for signal generation by GABA_BR1/GABA_BR2.

We studied the ligand-binding properties of GABA_BR2 and GABA_BR1/GABA_BR2 in membranes from transiently transfected COS-7 cells. GABA_BR2 was not labelled by the antagonist ligand ³H-

labelled CGP54626 or by agonists (³H-labelled GABA, ³H-labelled baclofen). The GABA_BR1/GABA_BR2 combination was labelled by ³H-labelled CGP54626 ($K_i = 0.8 \pm 0.02$ nM; $n = 3$), but not by the agonist radioligands. Overall, the binding properties of co-expressed GABA_BR1/GABA_BR2 were indistinguishable from those of GABA_BR1 alone, with respect to agonists (GABA, baclofen and 3-APMPA) and antagonists (phaclofen, CGP54626, CGP55845 and SCH50911) (data not shown). The absence of detectable high-affinity agonist binding to GABA_BR1/GABA_BR2, as well as to GABA_BR1 (ref. 2), is a notable distinction from the agonist/antagonist-binding profile of GABA_B in the central nervous system⁹ and may reflect limiting amounts of a specific G protein in the cell line.

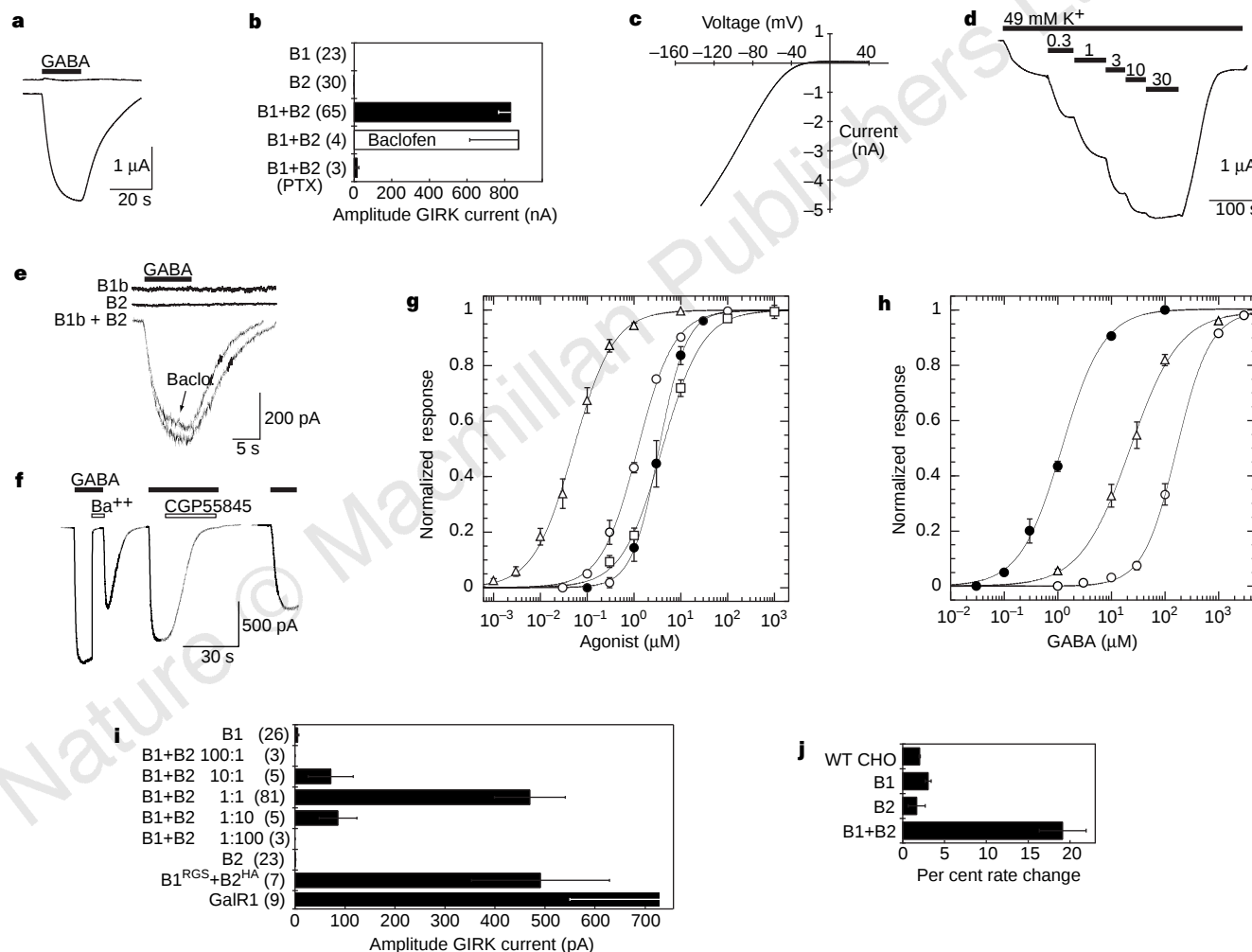


Figure 2 Signalling by co-expressed GABA_BR1 and GABA_BR2. **a**, Lower trace, response to GABA (100 μ M) of an oocyte expressing GABA_BR1, GABA_BR2 and GIRKs. Upper trace, a similar oocyte pretreated 6 h earlier with pertussis toxin (2 ng injected). **b**, Mean response amplitudes from oocytes expressing combinations of GABA_BR1 (B1) and GABA_BR2 (B2), plus GIRKs. Responses are to 100 μ M GABA (black bars) or 100 μ M baclofen (white bar). Numbers of observations are in parentheses. PTX, pertussis toxin. **c**, Voltage dependence of a GIRK current activated by 100 μ M GABA in an oocyte expressing GABA_BR1, GABA_BR2 and GIRKs. Ba²⁺-insensitive currents were subtracted. **d**, Concentration dependence of a GABA current recorded from a similar oocyte. Numbers over bars indicate the concentration of GABA (in μ M). **e**, Response to GABA or baclofen (each 100 μ M in 25 mM K⁺) in HEK293 cells expressing GIRKs and GABA_BR1b, GABA_BR2 or both. **f**, A HEK293 cell expressing GABA_BR1b, GABA_BR2 and GIRKs generates inward currents in response to GABA (filled horizontal bars, 100 μ M in 25 mM K⁺) that are blocked by 100 μ M Ba²⁺ or 1 μ M of the selective

antagonist CGP55845. The break in the trace represents 1 min. **g**, Agonist concentration-effect curves for 3-APMPA in oocytes (open triangles), GABA in oocytes (open circles) and in HEK293 cells (filled circles), and baclofen in oocytes (open squares). **h**, Shifts in the GABA concentration-response curve (filled circles) caused by 50 nM CGP55845 (open triangles) and 5 μ M CGP54626 (open circles). Each point is the average response from 4–6 oocytes $\pm$ s.e.m. **i**, Summary of mean response amplitudes from HEK293 cells co-transfected with various combinations and ratios of cDNAs. Responses are to 100 μ M GABA except for GalR1 where 10 nM galanin was used. To prepare different ratios of GABA_BR1:GABA_BR2 the most abundant cDNA was held constant at 0.6 μ g per dish and the other cDNA was reduced by a factor of 10 or 100. Numbers of observations are shown in parentheses. **j**, Microphysiometric response to baclofen (100 μ M) of CHO cells expressing combinations of GABA_BR1 and GABA_BR2 ($n = 4$). WT, wild-type.

The molecular mechanism by which protein co-expression confers functional activity is unknown. Varying the ratios of GABA_BR1 and GABA_BR2 cDNAs from 1:100 to 100:1 in HEK293 cells resulted in a symmetrical fall-off in response amplitude (Fig. 2i). This indicates that a 1:1 protein stoichiometry may be critical, and led us to postulate that the polypeptides form a heteromeric association. Biochemical evidence supports the idea that certain GPCRs can exist as homodimers^{3,4,6,7,22}, but the functional significance of this has been largely unstudied^{13,23}. We investigated the possibility of a physical association between GABA_BR1 and GABA_BR2 by using epitope-tagged versions of GABA_BR1 (RGS6His tag) and GABA_BR2 (haemagglutinin (HA) tag). C-terminal modification did not appear to alter the functional response: maximal current amplitudes (Fig. 2i) and EC₅₀ values for GABA (4.97 μ M, $n = 5$) were unchanged compared with HEK293 cells expressing the wild-type GABA_BR1/GABA_BR2 combination (3.42 μ M, $n = 5$). We studied the subcellular distribution of the epitope-tagged proteins in transfected cells by fluorescence microscopy. When expressed individually, GABA_BR1 and GABA_BR2 were localized in the vicinity of the plasma membrane as well as in the cytoplasm (data not shown). Optical sectioning of antibody-labelled cells by confocal microscopy confirmed the membrane-localization pattern; less labelling was present in the cytoplasm than at the plasma membrane and no labelling was detected in the nucleus. In co-transfected cells, there was a striking overlap in the distribution of the two epitope tags (Fig. 3a–c). Both proteins were prominently expressed near the plasma membrane. Furthermore, co-localization occurred within the cytoplasm, indicating that GABA_BR1 and GABA_BR2 may assemble in the endoplasmic reticulum. In contrast, the cellular distribution of an unrelated GPCR, neuropeptide Y Y5, differed considerably from that of GABA_BR2 (Fig. 3d), suggesting specificity in the association of GABA_BR2 with GABA_BR1.

Western blots of whole-cell extracts from cells expressing RGS6His–GABA_BR1, HA–GABA_BR2 or both proteins exhibited bands close to the predicted relative molecular masses (M_r values) of

the two proteins (GABA_BR1, 92,000; GABA_BR2, 97,000) and further bands corresponding to the predicted relative molecular masses for dimers (Fig. 4a, b). To determine whether GABA_BR1 and GABA_BR2 associate with each other in a heteromeric complex, we performed immunoprecipitations using membrane fractions enriched in plasma membrane as determined by the presence of Na⁺/K⁺ ATPase (Fig. 4c). In co-transfected cells only, HA–GABA_BR2 was detected in material immunoprecipitated using antibodies specific for the RGS6His–GABA_BR1 protein (Fig. 4d). This result confirms that both GABA_BR1 and GABA_BR2 are correctly targeted to the plasma membrane of HEK293 cells, and that the two proteins exist there in a heteromeric complex, perhaps as dimers.

Our data indicate that GABA_BR1 and GABA_BR2 may associate as subunits to produce a single pharmacologically and functionally defined receptor *in vivo*. Consistent with this view, *in situ* hybridization experiments provide evidence that GABA_BR1 and GABA_BR2 mRNAs co-localize within the same subpopulations of neurons, including Purkinje cells in the cerebellar cortex (Fig. 1d, e) and individual hippocampal pyramidal cells (Fig. 1f, g). On the other hand, the low level of expression of GABA_BR2 mRNA relative to GABA_BR1 mRNA in caudate putamen and hypothalamus (Fig. 1b, c) raises the possibility that other GABA_B-receptor homologues may associate with GABA_BR1 to produce new receptor subtypes in these regions.

The recent cloning of the genes encoding a family of accessory proteins that modify the binding and functional properties of a calcitonin-receptor-like receptor²⁴ shows that some seven-transmembrane-domain proteins require additional, unrelated proteins to reconstitute native GPCR activity. GABA_BR1 and GABA_BR2, to our knowledge, are the first examples of seven-transmembrane-domain proteins for which activity is dependent on an interaction with another member within the same gene family. There will be considerable interest in whether other GPCRs consist of heteromeric complexes. Many members of the superfamily of GPCRs, such as dopamine D₃ receptors, serotonin 5-HT₅ receptors and olfactory receptors, do not

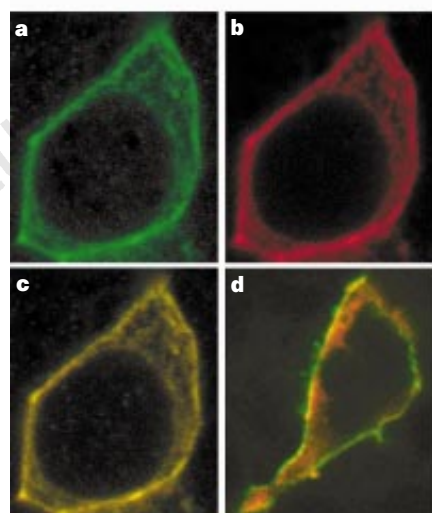


Figure 3 Co-localization of GABA_BR1 and GABA_BR2 in HEK293 cells by dual-wavelength scanning confocal microscopy. **a**, Green channel showing RGS6His–GABA_BR1 (labelled with FITC) in a cell expressing both RGS6His–GABA_BR1 and HA–GABA_BR2. **b**, Red channel showing HA–GABA_BR2 (labelled with TRITC) localization in the same cell. **c**, Dual-channel image of the same cell reveals a predominant yellow hue caused by the co-localization of the fluorescent tags for RGS6His–GABA_BR1 and HA–GABA_BR2. **d**, Dual-wavelength image of a cell expressing HA–GABA_BR2 (red) and Flag–neuropeptide-Y Y5 (green). Note the low degree of spatial overlap of the two receptors.

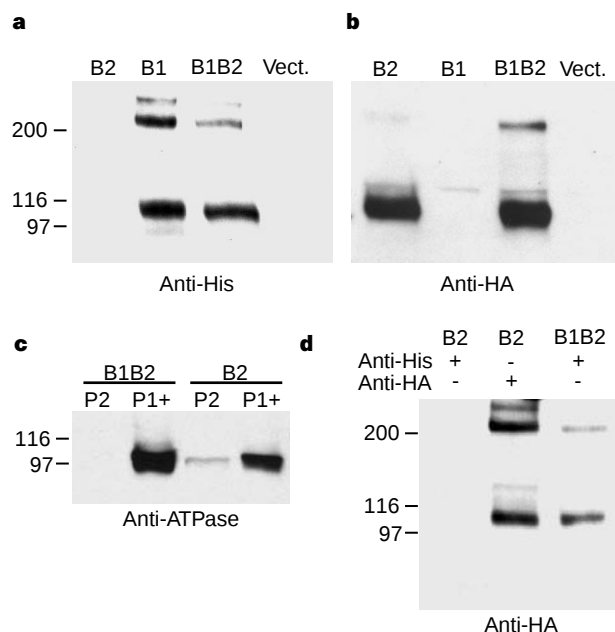


Figure 4 Identification of GABA_BR1 and GABA_BR2 in cell lysates and immunoprecipitates. Detection of **a**, RGS6His–GABA_BR1(B1) and **b**, HA–GABA_BR2(B2) in whole-cell extracts from cells expressing either or both tagged proteins. Both proteins appear to migrate as both monomeric and dimeric forms. **c**, Detection of Na⁺/K⁺ ATPase by anti- α 1-subunit antibodies in membrane fractions enriched in (P1+) or depleted of (P2) plasma membranes (50 μ g total protein per lane). **d**, Co-immunoprecipitation of RGS6His–GABA_BR1 (with anti-His antibody) and HA–GABA_BR2 from solubilized P1+ membrane fractions from cells expressing both proteins.

function well in heterologous expression systems and may require related partners²⁵ to generate native receptor activity. The growing list of receptors that may exist as homodimers^{3–7} points to the likelihood that both homomeric and heteromeric assemblies are more widespread among GPCRs than previously thought. □

Methods

Cloning of GABA_BR2 and GABA_BR1 cDNAs. A BLAST search of GenBank using the rat GABA_BR1a cDNA sequence identified two unannotated EST entries (accession numbers T07621 and Z43654) that had significant homology to this gene. We used primers corresponding to these EST sequences to isolate from a human hippocampal cDNA library a partial clone that comprised both of the EST sequences. Primers corresponding to transmembrane domains I and VII of the partial human clone were used at reduced stringency to amplify a fragment from rat hippocampus that showed 90% nucleotide identity to the human clone. Further primers corresponding to this rat cDNA fragment were used to identify a full-length clone, GABA_BR2, from a rat hypothalamic cDNA library²⁶. The GABA_BR2 clone was sequenced on both strands. Rat GABA_BR1b (referred to throughout the text as GABA_BR1) cDNA was obtained by screening a rat hypothalamic cDNA library²⁶ with polymerase chain reaction (PCR) primers corresponding to the published sequence². A rat GABA_BR1a clone was obtained by ligating a restriction fragment of GABA_BR1b, which contained the region common to the GABA_BR1 gene, to a PCR product containing the GABA_BR1a-specific 5' end.

In situ hybridization. Coronal sections (11 µm) of male Sprague–Dawley rat brain were prepared as described²⁷ with modifications to hybridization conditions. Sense and antisense oligonucleotides corresponding to positions 354–398 of GABA_BR2 and 1,076–1,120 of GABA_BR1b were labelled with ³⁵S-labelled dATP to a specific activity of 10⁹ d.p.m. µg⁻¹.

Electrophysiology. Oocytes were prepared and injected with synthetic mRNA encoding GIRK1 (refs 14, 16) and GIRK4 (ref. 15) as described^{26,28}. Oocytes were voltage-clamped at –80 mV. Drugs were applied by gravity in ND96 (ref. 28) or raised K⁺ (ref. 26) buffers. Concentration–response data were fitted with curves as described²⁶.

HEK-293 cells were co-transfected (Superfect, Qiagen) with various combinations of 0.6 µg each of the following: GABA_BR1b, GABA_BR2, GIRK1, GIRK4 and green fluorescent protein (GFP; pGreen Lantern, Gibco). Cells were voltage-clamped in whole-cell mode to a holding potential of –70 mV. Solutions used were (in mM): 125 KMeSO₄, 5 KCl, 5 NaCl, 2 MgCl₂, 11 EGTA, 10 HEPES, 1.0 Mg-ATP, 0.2 Na₂-GTP, pH 7.4, for the pipette and 130 NaCl, 4 KCl, 2 CaCl₂, 2 MgCl₂, 10 glucose, 10 sucrose, 10 HEPES, pH 7.4, for the bath. GIRK currents were recorded in a raised [K⁺] solution containing 25 mM K⁺ and a correspondingly lower concentration of Na⁺. Average values are shown as mean ± s.e.m.

Ligand binding. Competition binding experiments were performed as described²⁶ using membranes from COS-cells that had been transfected with plasmids encoding GABA_BR1 or GABA_BR2 or with both plasmids. Membranes were suspended at a concentration of ~10 µg ml⁻¹ in 50 mM Tris buffer with 2.5 mM CaCl₂ (pH 7.4) and incubated with 1.0 nM ³H-labelled CGP54626 for 60 min in the absence or presence of competitors. Incubations were terminated by rapid filtration through Whatman GF/C filters, and binding was quantified on a scintillation counter.

Microphysiometry. GABA_BR1, GABA_BR2 or both receptors were transiently expressed in CHO-K1 cells (LipofectAMINE, Gibco) and maintained in Ham's F-12 medium with 10% bovine serum. Cells were prepared for microphysiometric recording as described²⁹. Acidification rates reported are expressed as the per cent increase of the peak response over the baseline rate observed before drug application.

Construction of epitope-tagged receptors and confocal microscopy. Incorporation of sequences encoding the RGS6His or influenza virus haemagglutinin (HA) epitopes into the GABA_BR1 and GABA_BR2 genes, respectively, was performed by PCR. Each epitope was positioned immediately before the stop codon in the appropriate gene, and the constructs were subcloned into pcDNA3.1. Forty-eight hours after transfection, HEK293 cells were fixed for 20 min in 4% paraformaldehyde in PBS, permeabilized with 0.1% Triton-X100 and incubated with primary antibody for 1.5 h. Mouse monoclonal anti-RGS (Qiagen) and mouse anti-Flag (Boehringer) antibodies were labelled with fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse antibodies. Rat monoclonal anti-HA antibody (Boehringer) was

visualized with tetramethyl rhodamine isothiocyanate (TRITC)-conjugated rabbit anti-rat antibodies. Fluorescent images were obtained with a Zeiss LSM 410 confocal microscope using a ×100 oil-immersion objective.

Western blotting and immunoprecipitation. Western analysis of epitope-tagged proteins was done using whole-cell lysates of HEK293 cells collected 48 h after transfection in solution containing (in mM): 50 Tris/Cl, pH 7.4, 300 NaCl, 1.5 MgCl₂, 1 CaCl₂, protease inhibitors (Boehringer), 1% Triton-X100 and 10% glycerol. We subjected 20 µg total protein per lane to reducing SDS–PAGE, followed by blotting with either anti-HA or anti-4His antibody and then with sheep anti-rat or goat anti-mouse horseradish peroxidase (HRP)-linked secondary antibodies, respectively. Proteins were visualized with enhanced chemiluminescent substrates (Pierce).

Material for immunoprecipitations was obtained by sucrose gradient fractionation³⁰ of the P1 pellet obtained after mechanical disruption of transfected HEK293 cells. To confirm the enrichment of plasma membrane in the resulting P1⁺ pellet, we quantified Na⁺/K⁺ ATPase in the P1⁺ and P2 (primarily microsomal and vesicular³⁰) fractions by fluorescence detection of the anti-α1-subunit antibody (Research Diagnostics, clone 9A-5) on a phosphorimager (Molecular Dynamics). ATPase in the P1⁺ fractions used for immunoprecipitations was enriched by >50-fold compared with P2 fractions. We incubated 1–2 mg of plasma-membrane protein overnight at 4 °C with either 0.5 µg rat monoclonal anti-HA antibody or 0.5 µg mouse monoclonal anti-4His antibody (Qiagen). Immune complexes were bound to 20 µl protein A–agarose and isolated by repeated centrifugation and resuspension in buffer containing Triton-X100. Samples were analysed on western blots as described above.

Received 21 July; accepted 13 November 1998.

- Bowery, N. G. GABA_B receptor pharmacology. *Annu. Rev. Pharmacol. Toxicol.* **33**, 109–147 (1993).
- Kaupmann, K. *et al.* Expression cloning of GABA_B receptors uncovers similarity to metabotropic glutamate receptors. *Nature* **386**, 239–246 (1997).
- Hebert, T. E. *et al.* A peptide derived from a beta2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation. *J. Biol. Chem.* **271**, 16384–16392 (1996).
- Cvejic, S. & Devi, L. A. Dimerization of the delta opioid receptor: implication for a role in receptor internalization. *J. Biol. Chem.* **272**, 26959–26964 (1997).
- Maggio, R., Barbier, P., Fornai, F. & Corsini, G. U. Functional role of the third cytoplasmic loop in muscarinic receptor dimerization. *J. Biol. Chem.* **271**, 31055–31060 (1996).
- Ciruela, F. *et al.* Immunological identification of A1 adenosine receptors in brain cortex. *J. Neurosci. Res.* **42**, 818–828 (1995).
- Romano, C., Yang, W. L. & O'Malley, K. L. Metabotropic glutamate receptor 5 is a disulfide-linked dimer. *J. Biol. Chem.* **271**, 28612–28616 (1996).
- O'Hara, P. J. *et al.* The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. *Neuron* **11**, 41–52 (1998).
- Bowery, N. G., Hudson, A. L. & Price, G. W. GABA_A and GABA_B receptor site distribution in the rat central nervous system. *Neuroscience* **20**, 365–383 (1987).
- North, R. A. Drug receptors and the inhibition of nerve cells. *Br. J. Pharmacol.* **98**, 13–23 (1989).
- Gahwiler, B. H. & Brown, D. A. GABA_B-receptor-activated K⁺ current in voltage-clamped CA3 pyramidal cells in hippocampal cultures. *Proc. Natl Acad. Sci. USA* **82**, 1558–1562 (1985).
- Andrade, R., Malenka, R. C. & Nicoll, R. A. A G protein couples serotonin and GABA_B receptors to the same channels in hippocampus. *Science* **234**, 1261–1265 (1986).
- Lüscher, C., Jan, L. Y., Stoffel, M., Malenka, R. C. & Nicoll, R. A. G protein-coupled inwardly rectifying K⁺ channels (GIRKs) mediate postsynaptic but not presynaptic transmitter actions in hippocampal neurons. *Neuron* **19**, 687–697 (1997).
- Dascal, N. *et al.* Atrial G protein-activated K⁺ channel: expression cloning and molecular properties. *Proc. Natl Acad. Sci. USA* **90**, 10235–10239 (1993).
- Krapivinsky, G. *et al.* The G-protein-gated atrial K channel I_{KACH} is a heteromultimer of two inwardly rectifying K channel proteins. *Nature* **374**, 135–141 (1995).
- Kubo, Y., Reuveny, E., Slesinger, P. A., Jan, Y. N. & Jan, L. Y. Primary structure and functional expression of a rat G-protein-coupled muscarinic potassium channel. *Nature* **364**, 802–806 (1993).
- Kaupmann, K. *et al.* Structure, pharmacology and chromosomal localization of GABA-B receptors. *Soc. Neurosci. Abstr.* **23**, 954 (1997).
- Habert-Ortoli, E., Amiranoff, B., Loquet, L., Laburthe, M. & Mayaux, J. F. Molecular cloning of a functional human galanin receptor. *Proc. Natl Acad. Sci. USA* **91**, 9780–9783 (1994).
- Bon, C. & Galvan, M. Electrophysiological actions of GABA_B agonists and antagonists in rat dorso-lateral septal neurones *in vitro*. *Br. J. Pharmacol.* **118**, 961–967 (1996).
- Seabrook, G. R., Howson, W. & Lacey, M. G. Electrophysiological characterization of potent agonists and antagonists at pre- and postsynaptic GABA_B receptors on neurones in rat brain slices. *Br. J. Pharmacol.* **101**, 949–957 (1990).
- Brugger, F., Wicki, U., Olpe, H. R., Froestl, W. & Mickel, S. The action of new potent GABA_B receptor antagonists in the hemisectioned spinal cord preparation of the rat. *Eur. J. Pharmacol.* **235**, 153–155 (1993).
- Avissar, S., Amitai, G. & Sokolovsky, M. Oligomeric structure of muscarinic receptors is shown by photoaffinity labeling: subunit assembly may explain high- and low-affinity agonist states. *Proc. Natl Acad. Sci. USA* **80**, 156–159 (1983).
- Wreggett, K. A. & Wells, J. W. Cooperativity manifest in the binding properties of purified cardiac muscarinic receptors. *J. Biol. Chem.* **270**, 22488–22499 (1995).
- McLatchie, L. M. *et al.* RAMPs regulate the transport and ligand specificity of the calcitonin-receptor-like receptor. *Nature* **393**, 333–339 (1998).
- Nimchinsky, E. A., Hof, P. R., Janssen, W. G. M., Morrison, J. H. & Schmauss, C. Expression of dopamine D3 receptor dimers and tetramers in brain and in transfected cells. *J. Biol. Chem.* **272**, 29228–29237 (1997).
- Smith, K. E. *et al.* Cloned human and rat galanin GALR3 receptors: pharmacology and activation of G-protein inwardly rectifying K⁺ channels. *J. Cell Biol.* **273**, 23321–23326 (1998).
- Durkin, M. M. *et al.* Localization of messenger RNAs encoding three GABA transporters in rat brain: an *in situ* hybridization study. *Brain Res. Mol. Brain Res.* **33**, 7–21 (1995).

28. Quick, M. W. & Lester, H. A. Methods for expression of excitability proteins in *Xenopus* oocytes. *Methods Neurosci.* 19, 261–279 (1994).
29. Salon, J. A. & Owicki, J. C. Real-time measurements of receptor activity: applications of micro-physiometric techniques to receptor biology. *Methods Neurosci.* 25, 201–233 (1995).
30. Graham, J. in *Centrifugation (The Practical Approach Series)* (ed. Rickwood, D.) 161–182 (IRL, Oxford, 1984).

Acknowledgements. We thank T. Swayne for use of the confocal microscopy facility at Columbia University; H.-Y. Chang, K. Ogozalek, J. Huang and Y. Wan for support with molecular biology; P. Vayse for cell biology facilities; C. Corray for advice on membrane fractionation; and S. Rabacchi for helpful hints on protein immunodetection.

Correspondence and request for materials should be addressed to K.A.J. (e-mail: kjones@synapticcorp.com). The cDNA sequence for GABA_BR2 has been deposited at GenBank under accession number AF074482.

Heterodimerization is required for the formation of a functional GABA_B receptor

Julia H. White*, Alan Wise*, Martin J. Main*, Andrew Green*, Neil J. Fraser*, Graham H. Disney*, Ashley A. Barnes*, Piers Emson†, Steven M. Foord* & Fiona H. Marshall*

* Receptor Systems, Molecular Pharmacology Unit, GlaxoWellcome, Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, UK
† Department of Neurobiology, Babraham Institute, Babraham, Cambridge CB2 4AT, UK

GABA (γ-aminobutyric acid) is the main inhibitory neurotransmitter in the mammalian central nervous system, where it exerts its effects through ionotropic (GABA_{A/C}) receptors to produce fast synaptic inhibition and metabotropic (GABA_B) receptors to produce slow, prolonged inhibitory signals. The gene encoding a

GABA_B receptor (GABA_BR1) has been cloned¹; however, when expressed in mammalian cells this receptor is retained as an immature glycoprotein on intracellular membranes² and exhibits low affinity for agonists compared with the endogenous receptor on brain membranes. Here we report the cloning of a complementary DNA encoding a new subtype of the GABA_B receptor (GABA_BR2), which we identified by mining expressed-sequence-tag databases. Yeast two-hybrid screening showed that this new GABA_BR2-receptor subtype forms heterodimers with GABA_BR1 through an interaction at their intracellular carboxy-terminal tails. Upon expression with GABA_BR2 in HEK293T cells, GABA_BR1 is terminally glycosylated and expressed at the cell surface. Co-expression of the two receptors produces a fully functional GABA_B receptor at the cell surface; this receptor binds GABA with a high affinity equivalent to that of the endogenous brain receptor. These results indicate that, *in vivo*, functional brain GABA_B receptors may be heterodimers composed of GABA_BR1 and GABA_BR2.

Using a combination of bioinformatics and conventional molecular biology techniques, we cloned human homologues of the two alternatively spliced rat cDNAs encoding GABA_BR1a and GABA_BR1b, which differ in their amino termini. The human homologues are 98% identical to the respective rat cDNAs. Database searches with the human GABA_BR1a cDNA sequence identified several expressed sequence tags (ESTs) which show homology to the human GABA_BR1 sequence. Each of these ESTs encodes a part of the same receptor and, using this information, a new receptor subtype, which we have called R2, was assembled. Like R1, the R2 receptor is a seven-transmembrane-domain protein with a signal sequence and an extended extracellular N terminus, but it also possesses an unusually long intracellular C-

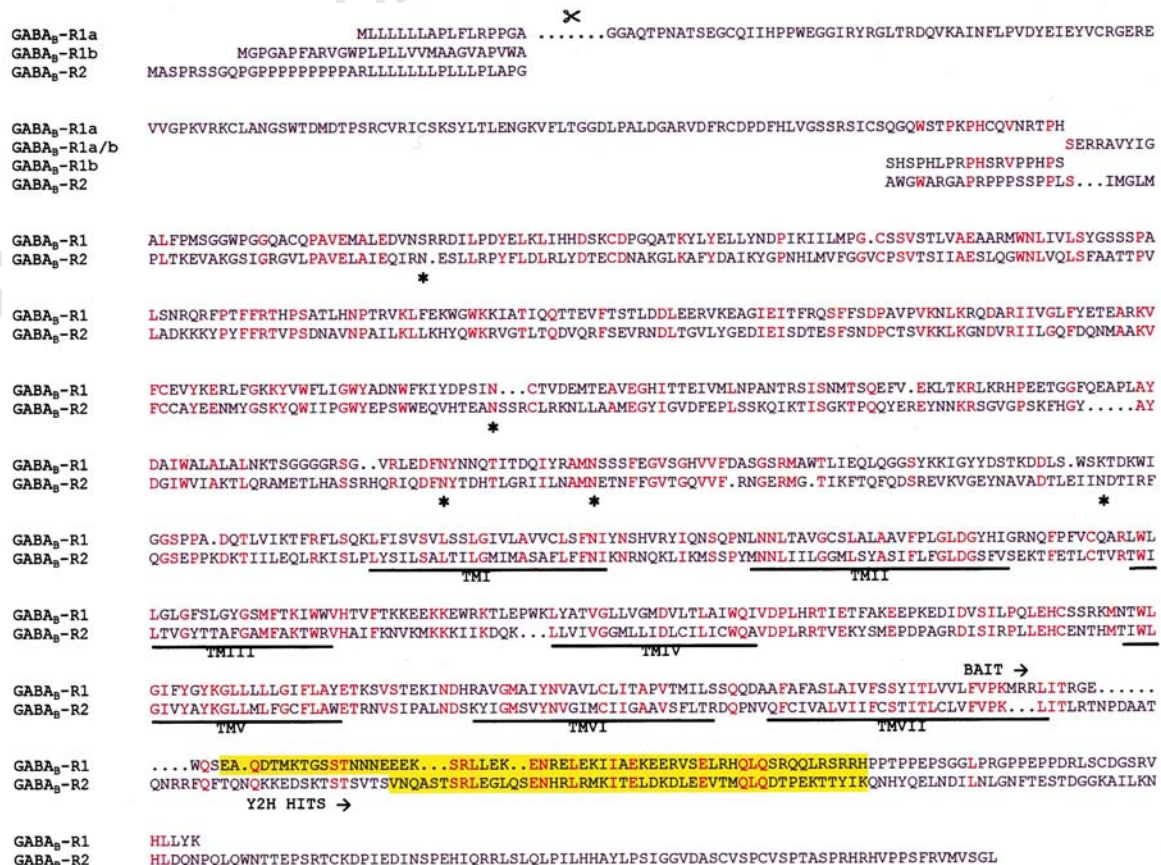


Figure 1 Amino-acid sequences of human GABA_BR1a, GABA_BR1b and GABA_BR2, aligned for comparison. The cleavage point of putative signal sequences (scissors), transmembrane (TM) domains and yeast two-hybrid 'bait' and hits (Y2H hits) are shown. Motif searches identified a coiled-coiled domain in the C terminus of both receptors (yellow) and five putative N-linked glycosylation sites (asterisks).

terminal tail (Fig. 1). At the protein level the two receptors are 35% identical and 54% similar.

Myc- or haemagglutinin (HA)-tagged and untagged constructs for human R1a, R1b or R2 were transiently transfected into HEK293T cells. We showed that each receptor is highly expressed by using gel electrophoresis of whole-cell membranes followed by western blotting. However, in assays of both [³⁵S]GTP-γS-binding and cyclic AMP production, we were unable to detect any functional activation of the cloned receptors by GABA or other GABA_B-selective agonists (data not shown), indicating that the receptors do not form a functional GABA_B receptor when expressed in mammalian cells.

A single transmembrane protein, receptor-activity-modifying protein 1 (RAMP1), is required for the trafficking, terminal glycosylation and expression of the calcitonin-receptor-like receptor as a functional receptor for calcitonin-gene-related peptides³. In addition, many receptors interact with synaptic 'anchoring proteins'^{4,5}. We and others², therefore, proposed that an accessory protein may be necessary for expression of maturely glycosylated GABA_BR1 at the cell surface. To investigate this possibility, we used the yeast two-hybrid system to search for proteins that interact with human GABA_BR1. We used the intracellular C terminus of GABA_BR1 as the 'bait' as it contains a 'coiled-coil' domain; such domains mediate several protein-protein interactions⁶. Screening of a human brain cDNA library (total of 4.3×10^6 cDNAs) produced several positive hits. A common hit was the C terminus of human GABA_BR2, which contains a coiled-coil domain similar to that seen in GABA_BR1 (Fig. 1). Further experiments in yeast confirmed that the interaction between the GABA_BR1 and GABA_BR2 C termini is specific, and showed that neither receptor will form homodimers (Fig. 2a). These results indicate that the two GABA_B receptors may

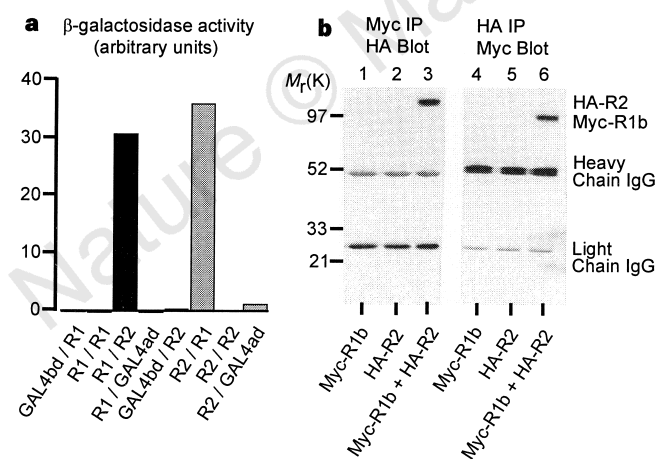


Figure 2 Interaction of GABA_BR1 and GABA_BR2. **a**, β -Galactosidase activity was measured in yeast strain Y190 expressing GABA_BR1 or GABA_BR2 C termini, either with empty vector or with each other in all combinations. Of each pair of proteins expressed, the first is the fusion with the Gal4 binding domain and the second is the Gal4ad fusion. (Gal4ad is Gal4 activation domain.) GAL4bd and GAL4ad denote empty-vector controls. β -Galactosidase activity is quantified relative to cell numbers. **b**, Co-immunoprecipitation studies. HEK293T cells were transfected with Myc-GABA_BR1b (lanes 1, 4), HA-GABA_BR2 (lanes 2, 5) or both receptors (lanes 3, 6), collected, lysed and immunoprecipitated (IP) using anti-Myc (lanes 1–3) or anti-HA (lanes 4–6) antisera. Immune complexes were subjected to SDS-PAGE and transferred to nitrocellulose, and Myc- or HA-tagged receptors were identified by immunoblotting with either anti-HA (lanes 1–3) or anti-Myc (lanes 4–6) antisera. In cells co-expressing Myc-GABA_BR1b and HA-GABA_BR2, the two receptors immunoprecipitated together, as indicated by immunoblotting with anti-HA (lane 3) or anti-Myc (lane 6) antisera.

heterodimerize through an interaction between their intracellular C-terminal tails, possibly through a conserved coiled-coil domain.

To test this hypothesis directly we carried out co-immunoprecipitation experiments in HEK293T cells expressing Myc-tagged GABA_BR1b and HA-tagged GABA_BR2 (Fig. 2b). Immunoprecipitation of GABA_BR1b from detergent-solubilized cell fractions with anti-Myc antisera led to detection of HA-GABA_BR2 within immune complexes (Fig. 2b, lane 3). Conversely, Myc-GABA_BR1b was detected on a western blot following precipitation of GABA_BR2 with anti-HA antisera (Fig. 2b, lane 6). These data are compelling evidence for a physical association between GABA_BR1b and GABA_BR2 following their expression in mammalian cells. Results from preliminary *in situ* hybridization studies indicate that, in the human brain, the alternative transcripts GABA_BR1a and GABA_BR1b each co-localize with the GABA_BR2 transcript in the cortex and

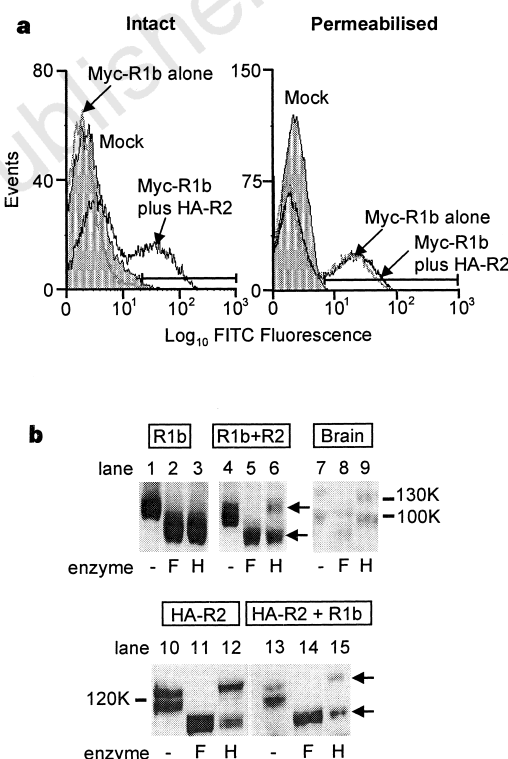


Figure 3 GABA_BR2 enables cell-surface expression of GABA_BR1b as a mature glycoprotein. **a**, Flow-cytometry analysis. In mock-transfected cells (shaded) and cells expressing Myc-GABA_BR1b alone, no anti-Myc FITC fluorescence was detected (marker indicates fluorescence measured over background) unless cells were permeabilized. Co-expression of HA-GABA_BR2 with Myc-GABA_BR1b led to surface expression of Myc-GABA_BR1b in 20% of cells sampled. **b**, Glycosylation status was assessed following treatment of membranes from HEK293T cells, transfected with GABA_BR1b (lanes 1–3), HA-GABA_BR2 (lanes 10–12) or combinations of GABA_BR1b, GABA_BR2 and HA-GABA_BR2 (lanes 4–6 and 13–15), with vehicle (–), endoglycosidase F or endoglycosidase H. Parallel experiments were carried out using rat brain membranes (lanes 7–9). Samples were resolved by SDS-PAGE and immunoblotted with antiserum 501 (lanes 1–9) or anti-HA antiserum (lanes 10–15). GABA_BR1 appears as a mature, endoglycosidase-H-resistant glycoprotein (top arrow, lane 6) only after co-expression with GABA_BR2. In contrast, endogenous rat GABA_BR1 (distinct bands at 130K (GABA_BR1a) and 100K (GABA_BR1b)) are expressed in an endoglycosidase-H-resistant form (lane 9). GABA_BR2 runs as a doublet at 120K, a component of which is endoglycosidase-H-resistant (top arrow). Top arrows represent terminally glycosylated forms of GABA_BR1 and GABA_BR2 in lanes 6 and 15, whereas bottom arrows represent core-glycosylated forms of GABA_BR1 and GABA_BR2.

hippocampus (data not shown), implying that a similar interaction occurs *in vivo*.

Further analysis of GABA_BR1 expressed in HEK293T cells showed that the receptor is retained as an immature glycoprotein on intracellular membranes (Fig. 3a; Fig. 3b, lanes 1–3), presumably explaining why we saw no responses of GABA_BR1 to GABA. We therefore proposed that co-expression of GABA_BR1 with GABA_BR2 may promote expression of GABA_BR1 at the cell surface and enable the GABA_BR1/GABA_BR2 heterodimer to function as a GABA_B receptor. We used flow-cytometry analysis to study the cellular distribution of immunotagged GABA_B receptors expressed in HEK293T cells (Fig. 3a). In intact cells expressing Myc–GABA_BR1b alone, no cell-surface anti-Myc immunoreactivity was detected. In contrast, immunoreactivity towards Myc was detected in 35% of permeabilized cells, reflecting intracellular expression of GABA_BR1b. However, when GABA_BR2 was co-transfected with Myc–GABA_BR1b, 20% of intact cells showed cell-surface anti-Myc immunoreactivity. As

the transfection efficiency in these experiments was ~35% (estimated from total expression in permeabilized cells), these data indicate that, in the presence of GABA_BR2, GABA_BR1b is efficiently moved to the cell membrane.

Endoglycosidases F and H can be used to differentiate between immature, core-glycosylated proteins and terminally glycosylated proteins that have passed through the Golgi apparatus^{3,7,8}. On a western blot, HEK293T cell membranes expressing GABA_BR1b produced a single distinct band of relative molecular mass 100,000 (M_r 100K; Fig. 3b, lane 1). This band was reduced to 80K following treatment with endoglycosidase F or H (Fig. 3b, lanes 2, 3), indicating that the expressed protein may be core-glycosylated but lacks terminal glycosylation. In contrast, following co-expression of GABA_BR1b and GABA_BR2 in HEK293T cells, a component of the GABA_BR1 immunoreactivity was resistant to endoglycosidase H (Fig. 3b, lane 6), indicating that a significant fraction of the GABA_BR1 receptor might now be mature glycoprotein. GABA_BR2 (which runs as a doublet at ~120K) was expressed at the cell surface and an endoglycosidase-H-insensitive band was observed even in the absence of GABA_BR1 (Fig. 3b, lane 12). Co-expression of GABA_BR1 had no effect on the glycosylation status of GABA_BR2 (Fig. 3b, lane 13–15).

We investigated whether the GABA_BR1/GABA_BR2 heterodimer behaves in a similar manner to the endogenous rat brain receptor when expressed in HEK293T cells. Expression of GABA_BR1 alone, but not GABA_BR2 alone, produced high levels of specific binding of the GABA_B antagonist ³H-labelled CGP54626. However, as reported for binding of ¹²⁵I-labelled CGP64213 (a structural analogue of ³H-labelled CGP54626)¹, agonist inhibition curves were shifted about 30-fold to the right (half-maximal inhibitory concentration (IC_{50}) = 28.8 μ M) compared with binding to rat brain membranes (IC_{50} = 1.1 μ M; Fig. 4a). Transfection of an increasing ratio of GABA_BR2 to GABA_BR1 led to a progressive leftwards shift in the GABA competition curve (data not shown). At a ratio (0.25 μ g GABA_BR1:1 μ g GABA_BR2 DNA) that gave equal expression of the two receptor proteins, as determined by western blotting, an IC_{50} of 2.1 μ M was recorded; this is comparable to the IC_{50} of 1.1 μ M determined with rat brain membranes (Fig. 4a). Co-expression of GABA_BR1a and GABA_BR2 (in combination with exogenous $G\alpha_{o1}$, a G-protein α -subunit) in HEK293T cells also resulted in robust, agonist-dependent stimulation of [³⁵S]GTP- γ S binding (Fig. 4b). This was concentration-dependent, with a half-maximal effector concentration (EC_{50} ; 95 ± 11 μ M) similar to that seen in rat brain membranes (59 ± 4 μ M). Similar results were obtained from these cells using inhibition of forskolin-evoked cyclic AMP as a measurement ($38 \pm 9\%$ inhibition, EC_{50} = 3.7 ± 1 μ M; data not shown). Finally, we co-expressed GABA_BR1a and GABA_BR2 in *Xenopus* oocytes to study coupling of GABA_B receptors to the G-protein-regulated potassium channels GIRK1 and GIRK4 (ref. 9). As seen in our other functional experiments, there were no responses to GABA in oocytes expressing either receptor in isolation (Fig. 4c; n = 7 for GABA_BR1a and n = 8 for GABA_BR2). However, in 21 out of 21 oocytes co-expressing GABA_BR1a and GABA_BR2, application of 100 μ M GABA or 1 mM baclofen activated a large inward current (343 ± 76 nA for GABA; Fig. 4c), which gave a current–voltage curve consistent with activation of the GIRK1/GIRK4 heteromer (Fig. 4d). In all these experiments, similar results were obtained with the GABA_BR1a and GABA_BR1b splice variants.

Our results indicate that the human GABA_B receptor is composed of two distinct, seven-transmembrane-domain receptor subunits, which exist either as a heterodimer or possibly as a larger oligomer. These data are, to our knowledge, the first evidence for heterodimerization of a seven-transmembrane-domain receptor. Co-expression of GABA_BR1 and GABA_BR2 appears to be a prerequisite for maturation and transport of GABA_BR1 to the plasma membrane and results in high-affinity GABA binding and G-protein activation.

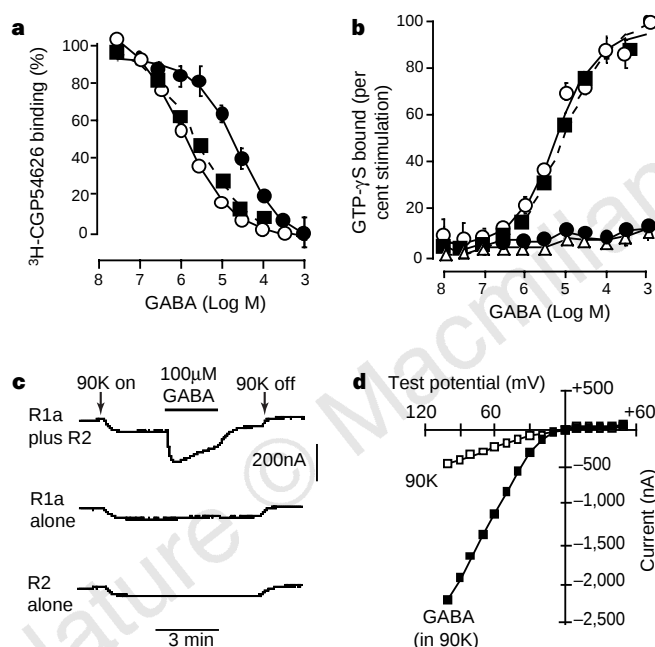


Figure 4 Ligand binding and functional studies. ³H-labelled CGP54626 competition binding and [³⁵S]GTP- γ S binding in membranes from HEK293T cells expressing GABA_BR1b (filled circles), GABA_BR2 (triangles), GABA_BR1b plus GABA_BR2 (squares, dashed line), or in rat brain membranes (open circles). **a**, A high level of specific binding of ³H-labelled CGP54626 was recorded in membranes expressing GABA_BR1b; binding was displaced in a concentration-dependent manner by GABA. Transfection of GABA_BR2 (which did not itself bind ³H-labelled CGP54626) with GABA_BR1 led to a leftwards shift in the GABA-displacement curve to a position similar to that in rat brain membranes. **b**, Stimulation by GABA of [³⁵S]GTP- γ S binding. Data are normalized to the peak response to GABA. No GABA-stimulated [³⁵S]GTP- γ S binding was recorded in GABA_BR1b- or GABA_BR2-expressing membranes; however, GABA stimulated a dose-dependent increase in [³⁵S]GTP- γ S binding both in rat brain and in cells expressing GABA_BR1b plus GABA_BR2. In all functional assays, identical results were obtained with the GABA_BR1a and GABA_BR1b splice variants. **c**, **d**, Oocyte data. Sample currents (**c**, top panel) and a current–voltage curve (**d**) are shown for individual oocytes expressing GABA_BR1a plus GABA_BR2 in combination with the potassium channels GIRK1 and GIRK4. Switching from low (ND96) to high (**c**, 90K; **d**, open squares) potassium solution led to an inward shift in holding current, indicating expression of the GIRK1 and GIRK4 potassium channels. Subsequent addition of 100 μ M GABA (filled squares in **d**) activated a large inward current through GIRK1/GIRK4.

Homodimerization of receptors has been reported for several seven-transmembrane-domain receptors¹⁰, including the δ -opioid¹¹, β_2 -adrenergic¹² and metabotropic glutamate (mGlu) receptors¹³. In the case of mGlu receptors, which are related to GABA_BR1 (ref. 1) and GABA_BR2, it has been suggested that homodimerization occurs through disulphide bridges in the N-terminal extracellular domains¹³. However, there are no cysteine residues in the corresponding regions of GABA_BR1 or GABA_BR2 and we believe that interaction occurs instead between the intracellular C termini of the two GABA_B receptors, probably through conserved coiled-coil domains, although other regions within the receptors might interact. The finding that the GABA_B receptor is a heterodimer containing GABA_BR1 and GABA_BR2 demonstrates a new mechanism of receptor behaviour and will further our understanding of the role of these receptors in the nervous system. Furthermore, the ability to reconstitute functional GABA_B receptors in recombinant systems will allow the development of new drugs at an important therapeutic target. □

Methods

Molecular biology. We used a bioinformatic approach to identify human ESTs showing homology to the rat GABA_BR1 receptor. These ESTs were aligned to provide coding sequences for human GABA_BR1a and GABA_BR1b (ESTs X90542, X90543, AA317417, D80024, AA348199, AA628887, T07518 and T06711) and for a putative GABA_B-receptor subtype, GABA_BR2. The predicted open reading frame of GABA_BR1a and GABA_BR1b was amplified by a combination of reverse transcription with polymerase chain reaction (RT-PCR) using human cerebellum poly(A)⁺ RNA (Clontech) and rapid amplification of cDNA ends (RACE)-PCR against Marathon-Ready human cerebellum cDNA (Clontech), and was cloned into vector pcDNA3.1 (Invitrogen). The entire open reading frame of GABA_BR2 could be represented by the human ESTs H14151, T07621, AA324303 and Z43654. Restriction fragments and PCR products generated from these clones were used to clone the full-length coding sequence into pcDNA3.1. For epitope-tagging of receptors, cDNAs encoding Myc (used with monoclonal antibody 9E10) and HA (used with monoclonal antibody 12CA5) epitopes were fused in-frame to the 5' end of cDNAs encoding GABA_BR1b and GABA_BR2. In each case the native signal sequence was removed and replaced with that for CD33 (ref. 14) (Myc-GABA_BR1b) or T8 (HA-GABA_BR1b, HA-GABA_BR2). All modified sequences were compared to 'native' sequences in the assays described and behaved in an identical manner.

Yeast two-hybrid studies. Yeast (*Saccharomyces cerevisiae* strain Y190)¹⁵ expressing a fusion protein containing a Gal4 binding domain and the GABA_BR1 C terminus were selected and transformed with a human brain 'Matchmaker' cDNA library (HL4004H, Clontech). Enough independent cDNAs were transformed to give a threefold representation of the library. Interacting clones were selected by growth under 20 mM 3-amino-1,2,4-triazole selection, followed by production of β -galactosidase as determined by a freeze-fracture assay. Plasmid DNA was recovered from yeast cells, transformed into *Escherichia coli* and sequenced.

Cell biology. HEK293T cells were maintained in DMEM medium containing 10% fetal calf serum and 2 mM glutamine, grown to 60–80% confluency in 60-mm dishes, transfected with cDNA using lipofectamine reagent (10 μ l, Life Technologies; total DNA = 3 μ g) and collected 48–72 h after transfection. Binding and glycosylation studies were done using plasma-membrane-containing P2 particulate fractions as described². Competition binding assays were done using displacement by GABA of the GABA_B antagonist ³H-labelled CGP54626 (ref. 16) (Tocris Cookson). Experiments were done in 50 mM Tris-HCl (pH 7.4) containing 2.5 mM Ca²⁺ and 40 μ M isoguvacine, and nonspecific binding was determined using 1 mM GABA. Bound ligand was recovered using a Brandel 48-well harvester and measured by liquid scintillation counting. [³⁵S]GTP- γ S-binding assays were done at room temperature in 96-well format as described¹⁷. Bound [³⁵S]GTP- γ S was determined by scintillation counting. In glycosylation studies, enzymatic removal of asparagine-linked (N-linked) carbohydrate moieties with endoglycosidases F and H was done essentially according to the manufacturer's instructions (Boehringer). GABA_B-receptor glycosylation status was studied following SDS-PAGE/

immunoblotting of samples and receptors were visualized by immunoblotting with antiserum 501, a polyclonal antibody raised against a synthetic peptide corresponding to the C-terminal 15 amino acids of GABA_BR1, or an anti-HA antiserum.

For co-immunoprecipitation experiments, HEK293T cells were collected and homogenized (in lysis buffer: 50 mM Tris-HCl, 150 mM NaCl, 1% (v/v) Nonidet P40, 0.5% (w/w) sodium deoxycholate, pH 7.5, supplemented with 'Complete' protease inhibitor cocktail tablets (1 tablet per 25 ml, Boehringer)) and precleared supernatant was collected by incubation with 50 μ l protein A-agarose (Boehringer). Immunoprecipitation with either anti-HA or anti-Myc antisera was allowed to proceed for 1 h at 4°C and capture of immune complexes was progressed overnight following the addition of 50 μ l protein A-agarose. Immune complexes were collected by microcentrifugation, released from protein A-agarose and analysed by SDS-PAGE followed by immunoblotting.

Flow-cytometry analysis was done using intact HEK293T cells, which were incubated with primary antibodies (9E10 or 12CA5) for 15 min at room temperature followed by secondary antibody (sheep anti-mouse Fab₂ coupled with fluorescein isothiocyanate (FITC)) for a similar period of time. For permeabilized cells, a Fix and Perm kit (Caltag) was used. Cells were analysed on a Coulter Elite flow cytometer set up to detect FITC fluorescence; 30,000 cells were analysed for each sample.

Oocytes. Capped *in vitro*-transcribed RNA (20–50 ng GABA_BR1a; GABA_BR2; GIRK1 and GIRK4 per oocyte) was injected into stage V–VI defolliculated oocytes¹⁸ and two microelectrode voltage clamp recordings were made 3–7 days after RNA injection from a holding potential of –60 mV. Oocytes were superfused with ND96 solution (96 mM NaCl, 2 mM KCl, 1 mM MgCl₂, 1.8 mM CaCl₂, 5 mM HEPES, pH 7.5, at 25°C) at a flow rate of 2 ml min^{–1}. To facilitate the recording of GIRK1/GIRK4 potassium currents, the extracellular solution was switched to a high-potassium solution (90 mM KCl, 1 mM MgCl₂, 1.8 mM CaCl₂, 5 mM HEPES). Recording electrodes had a resistance of 0.5–1.0 M Ω when filled with 3 M KCl. GABA was applied by addition to the superfusate.

Received 18 September; accepted 12 November 1998.

- Kaupmann, K. *et al.* Expression cloning of GABA_B receptors uncovers similarity to metabotropic glutamate receptors. *Nature* **386**, 239–246 (1997).
- Couve, A. *et al.* Intracellular retention of recombinant GABA_B receptors. *J. Biol. Chem.* **273**, 26361–26367 (1998).
- McLachie, L. M. *et al.* RAMPs regulate the transport and ligand specificity of the calcitonin-receptor-like receptor. *Nature* **393**, 333–339 (1998).
- Kornau, H. C., Schenker, L. T., Kennedy, M. B. & Seeburg, P. H. Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95. *Science* **269**, 1737–1740 (1995).
- Brakeman, P. R. *et al.* Homer: a protein that selectively binds metabotropic glutamate receptors. *Nature* **386**, 221–223 (1997).
- Lupas, A. Coiled coils: new structures and new functions. *Trends Biochem. Sci.* **21**, 375–382 (1996).
- Barsomian, G. D. *et al.* Cloning and expression of peptide-N₄-(N-acetyl-beta-D-glucosaminyl) asparagine amidase F in *Escherichia coli*. *J. Biol. Chem.* **265**, 6967–6972 (1990).
- Unson, C. G. *et al.* Characterization of deletion and truncation mutants of the rat glucagon receptor. Seven transmembrane segments are necessary for receptor transport to the plasma membrane and glucagon binding. *J. Biol. Chem.* **270**, 27720–27727 (1995).
- Kovoor, A., Nappey, V., Kieffer, B. L. & Chavkin, C. Mu and delta opioid receptors are differentially desensitized by the coexpression of β -adrenergic receptor kinase 2 and β -arrestin 2 in *Xenopus* oocytes. *J. Biol. Chem.* **272**, 27065–27611 (1997).
- Hebert, T. E. & Bouvier, M. Structural and functional aspects of G-protein-coupled receptor oligomerization. *Biochem. Cell. Biol.* **76**, 1–11 (1998).
- Cvejic, S. & Devi, L. A. Dimerization of the delta opioid receptor: implication for a role in receptor internalization. *J. Biol. Chem.* **272**, 26959–26964 (1997).
- Hebert, T. E. *et al.* A peptide derived from a beta2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation. *J. Biol. Chem.* **271**, 16384–16392 (1996).
- Romano, C., Yang, W.-L. & O'Malley, K. L. Metabotropic glutamate receptor 5 is a disulfide-linked dimer. *J. Biol. Chem.* **271**, 28612–28616 (1996).
- Simmons, D. & Seed, B. Isolation of a cDNA encoding CD33, a differentiation antigen of myeloid progenitor cells. *J. Immunol.* **141**, 2797–2800 (1988).
- Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**, 805–816 (1993).
- Bittiger, H., Reymann, N., Froest, W. & Mickel, S. J. ³H-CGP 54626: a potent antagonist radioligand for GABA_B receptors. *Pharmacol. Comm.* **2**, 23 (1992).
- Weiland, T. & Jakobs, K. H. Measurement of receptor-stimulated guanine 5'-O-(γ -thio) triphosphate in membrane preparations. *Methods Enzymol.* **237**, 3–13 (1994).
- Goldin, A. L. Maintenance of *Xenopus laevis* and oocyte injection. *Methods Enzymol.* **207**, 266–279 (1992).

Acknowledgements. We thank R. Solari and M. Roberts for comments on the manuscript, and the members of Receptor Systems for their support.

Correspondence and requests for materials should be addressed to F.H.M. (e-mail: fhm27375@glaxowellcome.co.uk). The cDNA sequences for human GABA_BR1a, GABA_BR1b and GABA_BR2 have been submitted to EMBL under the accession numbers AJ012185, AJ012186 and AJ012188, respectively.

GABA_B-receptor subtypes assemble into functional heteromeric complexes

Klemens Kaupmann*, Barbara Malitschek*, Valerie Schuler*, Jakob Heid*, Wolfgang Froestl*, Pascal Beck*, Johannes Mosbacher*, Serge Bischoff*, Akos Kulik†, Ryuichi Shigemoto†, Andreas Karschin‡ & Bernhard Bettler*

* Novartis Pharma AG, TA Nervous System, CH-4002 Basel, Switzerland

† Department of Morphological Brain Science, Faculty of Medicine, Kyoto University, Kyoto, Japan

‡ Molecular Neurobiology of Signal Transduction, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, D-37070 Göttingen, Germany

B-type receptors for the neurotransmitter GABA (γ -aminobutyric acid) inhibit neuronal activity through G-protein-coupled second-messenger systems, which regulate the release of neurotransmitters and the activity of ion channels and adenylyl cyclase¹. Physiological and biochemical studies show that there are differences in drug efficiencies at different GABA_B receptors, so it is expected that GABA_B-receptor (GABA_BR) subtypes exist². Two GABA_B-receptor splice variants have been cloned³ (GABA_BR1a

and GABA_BR1b), but native GABA_B receptors and recombinant receptors showed unexplained differences in agonist-binding potencies. Moreover, the activation of presumed effector ion channels in heterologous cells expressing the recombinant receptors proved difficult³⁻⁴. Here we describe a new GABA_B receptor subtype, GABA_BR2, which does not bind available GABA_B antagonists with measurable potency. GABA_BR1a, GABA_BR1b and GABA_BR2 alone do not activate Kir3-type potassium channels efficiently, but co-expression of these receptors yields a robust coupling to activation of Kir3 channels. We provide evidence for the assembly of heteromeric GABA_B receptors *in vivo* and show that GABA_BR2 and GABA_BR1a/b proteins immunoprecipitate and localize together at dendritic spines. The heteromeric receptor complexes exhibit a significant increase in agonist- and partial-agonist-binding potencies as compared with individual receptors and probably represent the predominant native GABA_B receptor. Heteromeric assembly among G-protein-coupled receptors has not, to our knowledge, been described before.

GenBank database searches revealed two expressed sequence tags, T07621 and HSC1HH041, with sequence similarity to the previously cloned GABA_BR1a/b receptors. We used this sequence information to isolate a 5.6-kilobase rat complementary DNA encoding a protein of 941 residues (Fig. 1a). Hydrophobicity analysis showed that this protein had a topological organization typical of a G-protein-coupled receptor (GPCR). The deduced protein sequence is most closely related to that of GABA_BR1a and

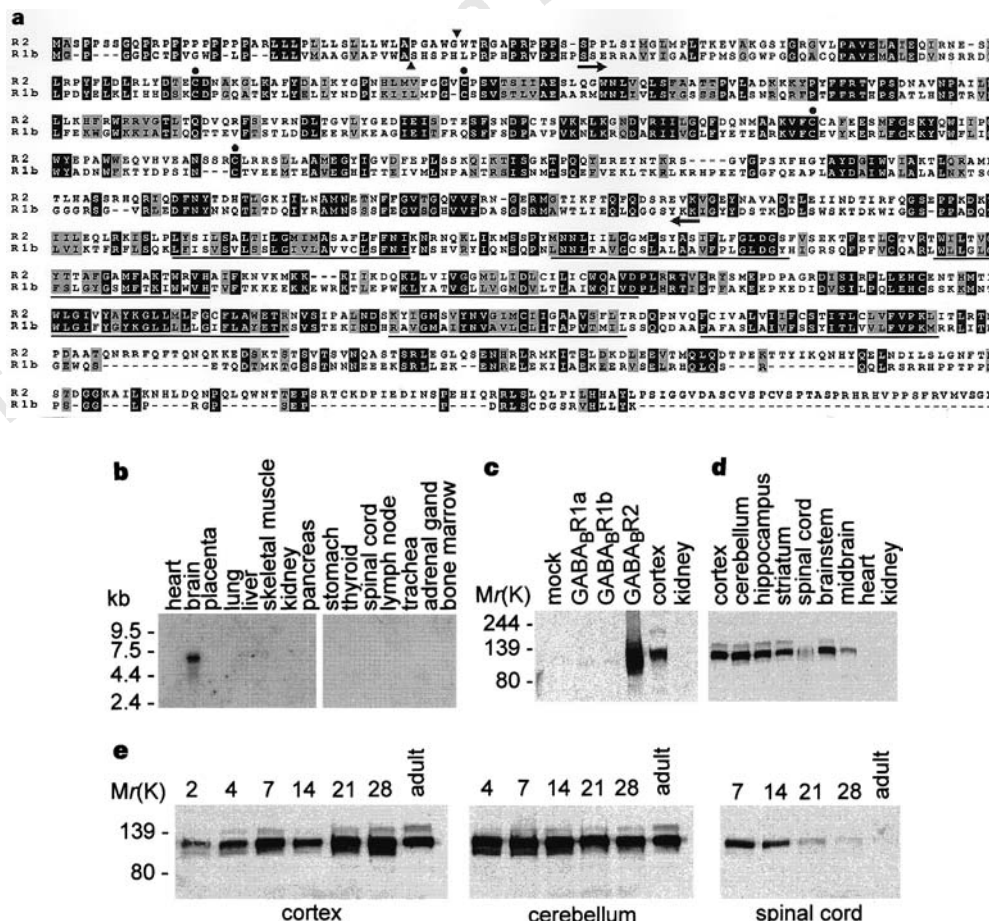


Figure 1 Sequence and expression of GABA_BR2 (R2). **a**, Alignment of rat GABA_BR2 and GABA_BR1b³ (R1b) amino-acid sequences. Identical amino acids are boxed in black; similar residues are shown in grey; membrane-spanning regions are underlined. Filled circles indicate conserved cysteine residues in the amino-terminal domain. Proposed sites of cleavage of signal peptides²⁴ are marked with arrowheads. The putative ligand-binding site, which shares homol-

ogy with the ligand-binding site of bacterial amino-acid-binding proteins²⁵, is delimited by arrows. **b**, Northern blot analysis of GABA_BR2 transcripts. **c**, Antibody AbC22 recognizes native and recombinant GABA_BR2 on immunoblots and does not crossreact with GABA_BR1a/b. **d**, Immunoblot analysis of GABA_BR2 protein in adult rats. **e**, Immunoblot analysis of GABA_BR2 in the cortex, cerebellum and spinal cord in rats of postnatal days 2 to 60 (adults).

GABA_BR1b (35% identity). Using northern blot analysis, we detected GABA_BR2 transcripts in the brain but not in the other tissues analysed (Fig. 1b). We raised antiserum AbC22, which is specific for the GABA_BR2 protein (Fig. 1c), and used it for immunoblot analysis; we detected a protein of relative molecular mass 110,000 (M_r 110K) in transfected cells that is expressed in all major brain areas (Fig. 1c, d). The GABA_BR2 protein is abundant in cortex and cerebellum throughout postnatal development whereas its expression in spinal cord gradually decreases (Fig. 1e).

We used membranes from transfected COS-1 cells expressing GABA_BR2 for binding assays with the GABA_B-receptor radioligands ¹²⁵I-labelled CGP64213 (0.5 nM), ³H-labelled CGP54626A (50 nM), ³H-labelled APPA (3-aminopropylphosphonic acid) (50 nM) and ³H-labelled GABA (50 nM) and for photoaffinity labelling experiments with ¹²⁵I-labelled CGP71872 (0.5 nM). None of the ligands bound to GABA_BR2 with measurable potency (data not shown). However, recombinant GABA_BR1a/GABA_BR1b exhibits a 100- to 150-fold reduced binding potency for agonists as compared with native receptors³, which precludes a direct measurement of binding to ³H-labelled agonists; therefore, a lack of detectable binding of ³H-labelled GABA to recombinant GABA_BR2 protein does not rule out the presence of an agonist-binding site in GABA_BR2. When GABA_BR1a or GABA_BR1b is expressed together with GABA_BR2, an up to tenfold increase in binding potency is observed, as measured by the inhibition of ¹²⁵I-labelled CGP64213 binding to the receptor by agonists and partial agonists (Fig. 2a, b). No further increase in binding potency occurred when GABA_BR2

cDNA was transfected in more than tenfold excess over GABA_BR1a/b cDNAs. In contrast, no significant change in the antagonist-binding pharmacology was detectable (Fig. 2c, d). There is evidence for the existence of dimeric receptors in the GABA_B receptor, metabotropic glutamate receptor and Ca²⁺-sensing receptor gene family^{5,6}. The conspicuous increase in agonist-binding potency observed here on co-expression of GABA_BR1a/b and GABA_BR2 led us to study the possibility of heteromeric GABA_B receptors.

GABA_B receptors generate late inhibitory postsynaptic potentials by activating Kir3-type K⁺ channels^{7,8}. Transfected HEK-293 cells that express Kir3 channels together with GABA_BR1a/b (ref. 4) or GABA_BR2 (Fig. 3a, b) infrequently yield a roughly twofold increase in K⁺ current (I_{kir}) after application of 50 μ M GABA. Only 1 out of 10 and 9 out of 62 cells expressing a I_{kir} of >200 pA (extracellular K⁺ concentration 25 mM) showed I_{kir} upregulation by GABA_BR1a/b and GABA_BR2, respectively. In GABA_BR2-expressing cells, I_{kir} upregulation is not blocked by the GABA_B antagonist CGP54626A (Fig. 3a), consistent with our failure to demonstrate binding of this antagonist to the receptor. This shows that GABA_BR2 is a GABA_B-receptor subtype with pharmacological properties that are different from those of GABA_BR1a/b.

Although 3 out of 40 *Xenopus* oocytes that were injected with both Kir3 and GABA_BR1a/b complementary RNAs showed small currents in response to α -baclofen (10 μ M) or GABA (100 μ M), we did not detect such responses in 28 oocytes expressing GABA_BR2 with Kir3 (Fig. 3c). However, when we co-injected cRNAs for GABA_BR2, GABA_BR1a and Kir3, application of α -baclofen

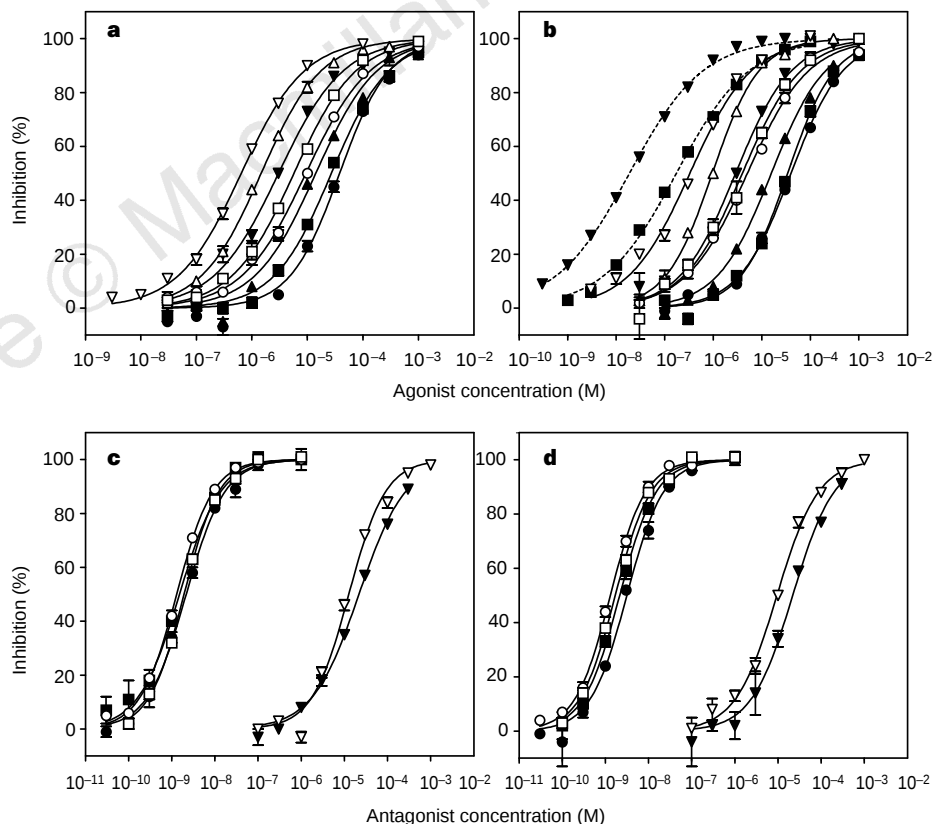


Figure 2 Binding pharmacology of GABA_BR2 co-expressed with GABA_BR1a or GABA_BR1b. The pharmacology of membranes from COS-1 cells expressing GABA_BR2 (20-fold excess) together with GABA_BR1a (**a, c**; open symbols) or GABA_BR1b (**b, d**; open symbols) was compared with the pharmacology of membranes from cells expressing GABA_BR1a (**a, c**; filled symbols) or GABA_BR1b individually (**b, d**; filled symbols) and with the pharmacology of brain GABA_B receptors (**b**, dotted lines, parallel assays). **a, b**, Inhibition of binding of ¹²⁵I-labelled CGP64213 to GABA_B receptors by agonists. **c, d**, Inhibition of binding of ¹²⁵I-labelled CGP64213 to GABA_B receptors antagonists. Immunoblotting and

photoaffinity labelling confirmed the expression of receptor protein (results not shown). The IC₅₀ values for agonists are shifted about 8- to 9-fold (GABA_BR1b/GABA_BR2) or 2.5- to 5-fold (GABA_BR1a/GABA_BR2) towards higher potencies in co-transfected cells. The partial agonist CGP47656 (ref. 3) shows a tenfold increase in potency. Antagonist affinities do not differ significantly. Results from typical experiments performed in triplicate are shown. In **a, b**, squares indicate that the agonist GABA was used; circles, α -baclofen; inverted triangles, APPA; and triangles, CGP47656. In **c, d**, circles indicate the use of the antagonist CGP54626A; squares, CGP64213; and inverted triangles, CGP35348.

(10 μ M), APPA (1 μ M) and GABA (10 μ M) evoked inward currents at -70 mV in 38 of 52 oocytes (Fig. 3c–e). On average, the upregulation of I_{kir} was $32 \pm 5\%$. The half-maximal effective concentration (EC_{50}) value for L-baclofen at GABA_BR1a/GABA_BR2 and GABA_BR1b/GABA_BR2 receptors expressed in oocytes is 5 μ M (Fig. 3e). This EC_{50} is similar to values obtained with GABA_BR1a (11.3 μ M)⁴ and GABA_BR1a/GABA_BR2 (1.5 μ M; see Supplementary Information) expressed in HEK-293 cells. It also approximately matches the values obtained for native GABA_B receptors^{7,9,10}. CGP54626A blocked I_{kir} upregulation with a half-maximal inhibitory concentration (IC_{50}) of 10 nM (Fig. 3f).

The requirement for two distinct GABA_B receptors for a robust coupling of receptor activation to I_{kir} explains in retrospect why it

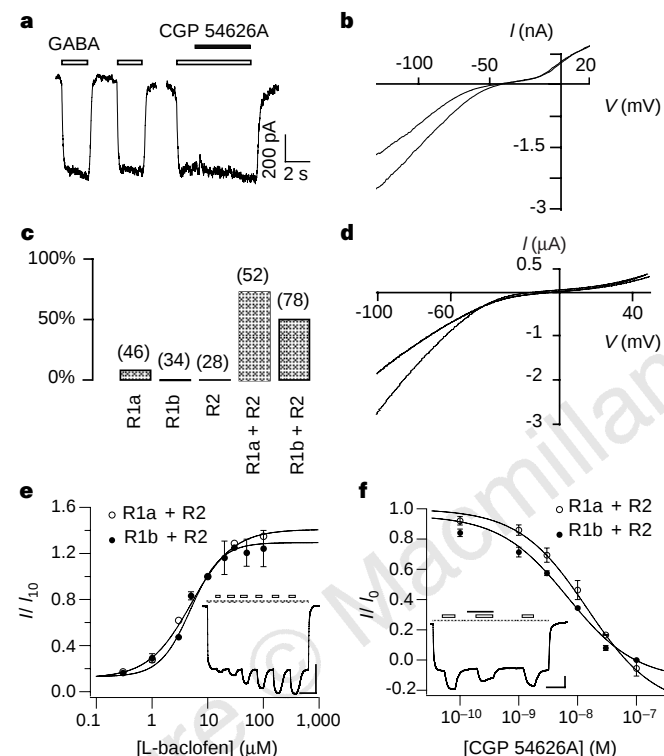


Figure 3 Coupling of GABA_B receptors to Kir3 channels in transfected HEK-293 cells and *Xenopus* oocytes. **a, b**, Current recordings from HEK-293 cells co-expressing GABA_BR2 (R2) and concatenated pairs of Kir3.1/3.2 channels in an extracellular K⁺ concentration of 25 mM. **a**, 50 μ M GABA and 1 μ M CGP54626A were applied as indicated. Holding potential was 120 mV. **b**, Current-voltage relation (I - V) in the absence (upper trace) and presence (lower trace) of 50 μ M GABA. **c-f**, Recordings from *Xenopus* oocytes expressing GABA_BR1a(R1a), GABA_BR1b (R1b) or GABA_BR2 (R2) or combinations thereof together with Kir3.1/3.2/3.4 channels. **c**, Percentage of oocytes that exhibit an upregulation of I_{kir} of $>2\%$ after superfusion of GABA (100 μ M) or L-baclofen (10 μ M). The numbers of cells tested is shown in parentheses. **d**, I - V in the absence (upper trace) and presence (lower trace) of L-baclofen (10 μ M). **e**, Concentration-response curve (CRC) for L-baclofen. Agonist-induced currents are normalized to those evoked by 10 μ M L-baclofen (I_{10}). The EC_{50} values for R1a + R2 and R1b + R2 are 5.0 ± 1.5 μ M ($n = 4$) and 4.9 ± 0.7 μ M ($n = 6$), respectively. Inset shows raw current trace of a CRC for R1a + R2; the time of application of high-potassium Ringer solution (dotted line) and of (open bars from left to right) 0.3, 1, 3, 10, 30 and 100 μ M L-baclofen are indicated. Scale bars, vertical, 500 nA and horizontal, 100 s. **f**, CRC for CGP54626A. Values are normalized to currents evoked by 10 μ M L-baclofen in the absence of antagonist (I_0). The IC_{50} values for R1a + R2 and R1b + R2 are 15 ± 6 nM ($n = 3$) and 7 ± 0.8 nM ($n = 3$), respectively. The inset shows the blockage of I_{kir} upregulation by 10 nM CGP54626A (black bar), 10 μ M L-baclofen (open bars), and high-potassium Ringer solution (dotted line). Scale bars, 200 nA and 100 s. **e, f**, The smooth lines represent Hill equations fitted to the data points.

has been possible to demonstrate modulation of I_{kir} in oocytes injected with cerebellar poly(A)⁺ RNA¹¹, but not in oocytes injected with GABA_BR1a/b cRNA³. Assembly of heteromeric receptors may be required for effective transport of the receptor to the membrane, as GABA_BR1a/b receptors by themselves fail to reach the cell surface in heterologous cells¹². The recombinant expression experiments therefore indicate that efficient GABA_B-receptor coupling *in vivo* may require assembly of heteromeric receptors. The increased potency of agonist binding observed with heteromeric GABA_B receptors (Fig. 2a, b) could then arise from a more efficient coupling to G proteins¹³.

We confirmed the existence of heteromeric GABA_B receptors in immunoprecipitation experiments (Fig. 4). Antibody AbC22, raised against GABA_BR2, efficiently co-precipitated the GABA_BR1a/b proteins, tagged with ¹²⁵I-labelled CGP71872, from cortical membrane preparations (Fig. 4a). Conversely, antibody Ab60696, which recognizes an epitope shared by the GABA_BR1a/b variants, co-precipitated the GABA_BR2 receptor (Fig. 4b). Immunoprecipitation experiments with transfected cells expressing receptor combinations showed that GABA_BR2 can associate with GABA_BR1a and GABA_BR1b individually (Fig. 4c). We did not find any evidence for the formation of intermolecular disulphide bonds between GABA_B receptors; such bonds do form within dimers of metabotropic glutamate receptor 5 (ref. 6) and the Ca²⁺-sensing receptor⁵.

In situ hybridization shows that there are high levels of GABA_BR2 messenger RNA in the cortex and hippocampus (Fig. 5a). Throughout the brain, the patterns of distribution of GABA_BR1a/b and GABA_BR2 mRNA largely overlap (Fig. 5a, b). GABA_BR2 transcript levels are low as compared with amounts of GABA_BR1a/b transcripts in the basal forebrain (striatum and nucleus accumbens) and the olfactory bulb. In the cerebellum, high levels of both GABA_BR1b (ref. 4) and GABA_BR2 transcripts are found in Purkinje cells, demonstrating co-expression in neuronal populations (Fig. 5c–f). Using immunoelectron microscopy, we determined whether GABA_BR1a/b and GABA_BR2 co-localize at neuronal membranes in the molecular layer of the cerebellum (Fig. 5g, h). Immunogold labelling of both receptors occurs at extrasynaptic sites of Purkinje cell spines that form type 1 synapses with parallel fibre terminals (Fig. 5g; shown for GABA_BR2). Double labelling showed that most

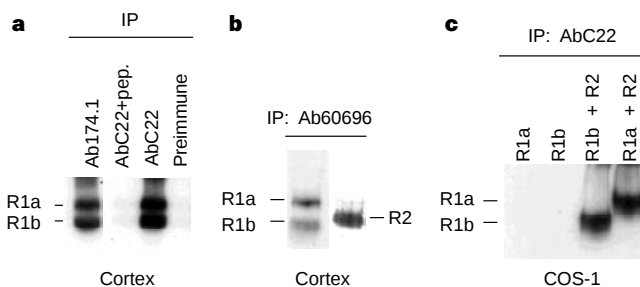


Figure 4 Immunoprecipitation analysis of GABA_B receptors. GABA_B-receptor expression in **a, b**, cortex and **c**, transfected COS-1 cells. Membranes were labelled with ¹²⁵I-labelled CGP71872, and proteins were immunoprecipitated (IP) with antibody AbC22, Ab60696 or Ab174.1 and subjected to SDS-PAGE. GABA_BR1 and GABA_BR1b proteins labelled with ¹²⁵I-labelled CGP71872 were detected using phosphorimaging. GABA_BR2 was detected on immunoblots using antibody AbC22. **a**, Ab174.1 directed against GABA_BR1 precipitates GABA_BR1a and GABA_BR1b labelled with ¹²⁵I-labelled CGP71872 (lane 1). AbC22 directed against GABA_BR2 co-precipitates ¹²⁵I-labelled GABA_BR1a and GABA_BR1b (lane 3). Immunoprecipitation was blocked by the corresponding antigenic peptide (10 μ g, lane 2) and preimmune serum (lane 4). **b**, Ab60696 directed against GABA_BR1a and GABA_BR1b co-immunoprecipitates GABA_BR2. **c**, GABA_BR1a and GABA_BR1b are co-immunoprecipitated with AbC22 when expressed in pairwise combination with GABA_BR2. AbC22 does not precipitate GABA_BR1a or GABA_BR1b in COS-1 cells that do not express GABA_BR2.

of the GABA_BR2-positive spines were also labelled for GABA_BR1a/b and vice versa, indicating extensive co-localization (Fig. 5h). Thus our data indicate that GABA_B receptors may form heteromers at select neuronal sites.

The fact that heterologous GABA_B receptors couple to Kir3 (shown here) and to adenylyl cyclase³, and that GABA_BR1a/b-containing receptors inhibit high-voltage-activated Ca²⁺ channels¹⁴, shows that all major effects of native GABA_B receptors could relate to those of the characterized heteromeric receptors. Given the precedent for dimerization among GPCRs^{5,6} we expect a heterodimeric configuration for the most abundant native GABA_B receptors, but another stoichiometry is possible. The coupling of individual GABA_BR1 (ref. 4) and GABA_BR2 (Fig. 3a, b) receptors to Kir3 may be the consequence of the expression of excessive amounts of receptors that force targeting to the cell surface. Nevertheless, the occurrence of homomeric receptors *in vivo* cannot be ruled out. The remaining tenfold discrepancy in apparent agonist-binding potency between heteromeric recombinant and native

receptors may be explained by receptor modification (for example, by phosphorylation) or by differences in the relative expression levels of G proteins and receptors¹⁵, and other proteins that assemble with GPCRs may alter the pharmacological properties of GABA_B receptors^{16,17}. □

Methods

Ligands. GABA_B ligands were synthesized in house. ¹²⁵I-labelled CGP64213 and ¹²⁵I-labelled CGP71872 were radiolabelled by ANAWA (Switzerland) to a specific activity of 2,000 Ci mmol⁻¹.

Cloning of GABA_BR2. Expressed sequence tags were cloned by reverse-transcription with polymerase chain reaction (RT-PCR) using cDNAs synthesized from human cerebellum poly(A)⁺ mRNA (Clontech). The PCR products were used to screen a rat brain cDNA library as described³.

In situ hybridization and northern blot analysis. *In situ* hybridization histochemistry using 10-μm cryosections (post-fixed with 4% paraformaldehyde) of rat brain (male Tif RAI f (specific pathogen free), weighing 250 g) was as described¹⁸. Riboprobes for GABA_BR2 were derived from a 724-base-pair *NheI/SacI* cDNA fragment subcloned into Bluescript (Stratagene). For GABA_BR1 a pan probe that is common to GABA_BR1a and GABA_BR1b was used³. Dipped slides were exposed for 15 days. Northern blots using poly(A)⁺ mRNA (Clontech) were hybridized as described³.

Cell culture and transfections. COS-1 and HEK-293 cells were obtained from the American Type Culture Collection and transfected with receptor cDNAs as described³.

Membrane preparations. Membranes from transfected cells were collected 3 days after cell transfection. Cell lysates of transfected cells and synaptic membranes from rat tissues (male Tif RAI f (SPF)) were prepared as described^{19,20}.

Photoaffinity labelling, antibodies and immunoprecipitation. Photolabelling with ¹²⁵I-labelled CGP71872 was done as described³. Antibodies Ab60696 (Pharmingen) and Ab174.1 are directed against carboxy-terminal epitopes of GABA_BR1a/b (ref. 19). To generate polyclonal antibody AbC22 in New Zealand white rabbits, we expressed a GABA_BR2 fragment corresponding to amino acids 806–907 as glutathione-S-transferase fusion protein in *Escherichia coli* using plasmid pGEX-2T (Pharmacia). For immunoprecipitation, proteins were solubilized in 50 mM HEPES, pH 7.5, 150 mM NaCl, 1.5 mM MgCl₂, 1 mM EGTA, 10% glycerol and 1% Triton-X100 including protease inhibitors (Boehringer) at 4°C for 1 h and centrifuged for 10 min at 10,000g; the supernatant was then centrifuged at 100,000g for 1 h. After twofold dilution in buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 10% glycerol, 0.1% Triton-X100, protease inhibitors), protein A–Sepharese pellets were washed twice with the above buffer. Immunoblotting was done as described¹⁹.

Pre-embedding immunoelectron microscopy. Cerebellar sections of Wistar rats were treated for single and double immunolabelling with anti-GABA_BR2 (guinea pig) and anti-GABA_BR1 (rabbit) antibodies as described²¹.

Electrophysiology. Whole-cell patch-clamp²² and oocyte²³ recordings were done as described. cRNAs were injected into oocytes at a concentration ratio of 4:1:0.2 (GABA_BR2:GABA_BR1a/b:Kir3.1/3.2/3.4). In high-potassium Ringer solution, 87.5 mM NaCl was replaced by KCl.

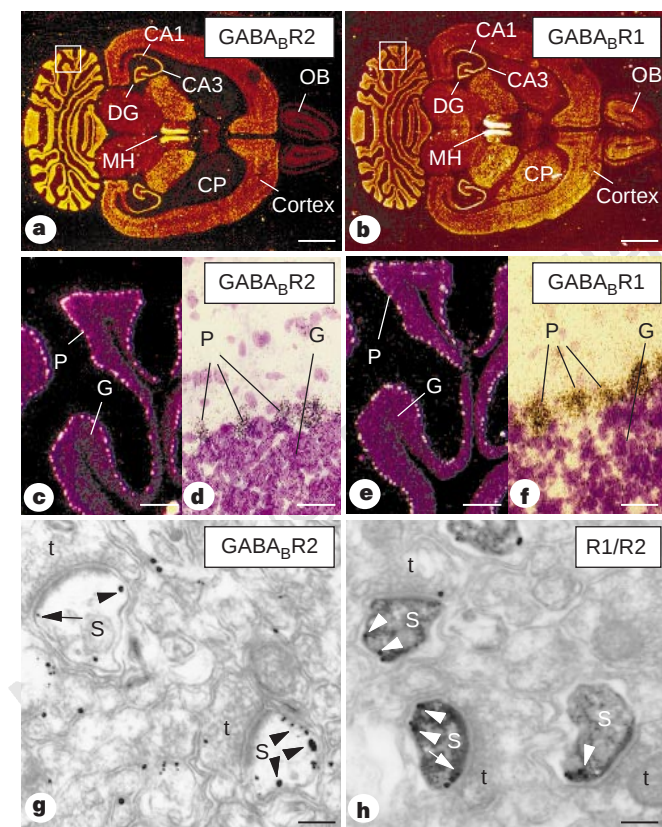


Figure 5 Cellular and subcellular distribution of GABA_B-receptor mRNA and protein in rat brain. **a–f**, *In situ* hybridization analysis. Autoradiograms of horizontal sections in dark-field (**a–c**, **e**) and bright-field (**d**, **f**) illuminations from the dorsal tier of the brain (**a**, **b**) and the lobules of the cerebellar cortex (**c–f**). Transcripts of both GABA_BR1 and GABA_BR2 are abundant in the medial habenula (MH), the pyramidal cells of the CA1–CA3 subfields of the hippocampus, the granular layer of the dentate gyrus (DG) and the cortex. GABA_BR2 transcripts are less abundant than GABA_BR1 transcripts in the caudate putamen (CP) and the olfactory bulb (OB). GABA_BR2 and GABA_BR1 transcript expression is higher in Purkinje cells (P) than in the granule layer (G). **g**, **h**, Electron micrographs in the molecular layer showing extensive co-localization of GABA_BR2 and GABA_BR1 protein in Purkinje cell spines (s) that form type 1 (asymmetrical) synapses with parallel fibre terminals (t): labelling is with immunogold for GABA_BR2 (**g**) and GABA_BR1 (**h**) and immunoperoxidase for GABA_BR2 (**h**). Immunoparticles are present at extrasynaptic sites (arrowheads) of the Purkinje cell plasma membrane. Some particles are found at the edge (arrows) of the postsynaptic density area. Scale bars. **a**, **b**, 1.3 mm; **c**, **e**, 120 μm; **d**, **f**, 10 μm; and **g**, **h**, 0.2 μm.

Received 25 September; accepted 4 November 1998.

- Kerr, D. I. & Ong, J. GABA_B receptors. *Pharmacol. Ther.* **67**, 187–246 (1995).
- Bettler, B., Kaupmann, K. & Bowery, N. G. GABA_B receptors: drugs meet clones. *Curr. Opin. Neurobiol.* **8**, 345–350 (1998).
- Kaupmann, K. *et al.* Expression cloning of GABA_B receptors uncovers similarity to metabotropic glutamate receptors. *Nature* **386**, 239–246 (1997).
- Kaupmann, K. *et al.* Human GABA_B receptors are differentially expressed and regulate inwardly rectifying K⁺ channels. *Proc. Natl Acad. Sci. USA* (in the press).
- Ward, D. T., Brown, E. M. & Harris, H. W. Disulfide bonds in the extracellular calcium-polyvalent cation-sensing receptor correlate with dimer formation and its response to divalent cations *in vitro*. *J. Biol. Chem.* **273**, 14476–14483 (1998).
- Romano, C., Yang, W.-L. & O'Malley, L. Metabotropic glutamate receptor 5 is a disulfide-linked dimer. *J. Biol. Chem.* **271**, 28612–28616 (1996).
- Sodickson, D. L. & Bean, B. P. GABA_B receptor-activated inwardly rectifying potassium current in dissociated hippocampal CA3 neurons. *J. Neurosci.* **16**, 6374–6385 (1996).
- Lüscher, C., Jan, L. Y., Stoffel, M., Malenka, R. C. & Nicoll, R. A. G protein-coupled inwardly rectifying K⁺ channels (GIRKs) mediate postsynaptic but not presynaptic transmitter actions in hippocampal neurons. *Neuron* **19**, 687–695 (1997).
- Slesinger, P. A., Stoffel, M., Jan, Y. N. & Jan, L. Y. Defective γ-aminobutyric acid type B receptor-activated inwardly rectifying K⁺ currents in cerebellar granule cells isolated from weaver and GIRK2 null mutant mice. *Proc. Natl Acad. Sci. USA* **94**, 12210–12217 (1997).
- Inoue, M., Matsuo, T. & Ogata, N. Characterization of pre- and postsynaptic actions of (–)-baclofen in the guinea-pig hippocampus *in vitro*. *Br. J. Pharmacol.* **84**, 843–851 (1985).

11. Uezono, Y. *et al.* Activation of inwardly rectifying K⁺ channels by GABA_B receptors expressed in *Xenopus* oocytes. *Neuroreport* **9**, 583–587 (1998).
12. Couve, A. *et al.* Intracellular retention of recombinant GABA_B receptors. *J. Biol. Chem.* **273**, 26361–26367 (1998).
13. Hill, D. R., Bowery, N. G. & Hudson, A. L. Inhibition of GABA_B receptor binding by guanyl nucleotides. *J. Neurochem.* **42**, 652–657 (1984).
14. Morris, S. J., Beatty, D. M. & Chronwall, B. M. GABA_B R1a/R1b-type receptor antisense deoxynucleotide treatment of melanotrophic blocks chronic GABA_B receptor inhibition of high-voltage-activated Ca²⁺ channels. *J. Neurochem.* **71**, 1329–1332 (1998).
15. Kenakin, T. Differences between natural and recombinant G-protein coupled receptor systems with varying receptor/G-protein stoichiometry. *Trends Pharmacol. Sci.* **18**, 456–464 (1997).
16. McLatchie, L. M. *et al.* RAMPs regulate the transport and ligand specificity of the calcitonin-receptor-like receptor. *Nature* **393**, 333–339 (1998).
17. Dwyer, N. D., Troemel, E. R., Sengupta, P. & Bargmann, C. I. Odorant receptor localization to olfactory cilia is mediated by ODR-4, a novel membrane-associated protein. *Cell* **93**, 455–466 (1998).
18. Bischoff, S., Barhanin, J., Bettler, B., Mulle, C. & Heinemann, S. Spatial distribution of kainate receptor subunit mRNA in the mouse basal ganglia and ventral mesencephalon. *J. Comp. Neurol.* **379**, 541–562 (1997).
19. Malitschek, B. *et al.* Developmental changes in agonist affinity at GABA_B R1 receptor variants in rat brain. *Mol. Cell. Neurosci.* **12**, 56–64 (1998).
20. Olpe, H.-R. *et al.* CGP 35348: a centrally active blocker of GABA_B receptors. *Eur. J. Pharmacol.* **187**, 27–38 (1990).
21. Shigemoto, R. *et al.* Target-cell-specific concentration of a metabotropic glutamate receptor in the presynaptic active zone. *Nature* **381**, 523–525 (1996).
22. Wischmeyer, E. *et al.* Subunit interactions in the assembly of neuronal Kir3.0 inwardly rectifying K⁺ channels. *Mol. Cell. Neurosci.* **9**, 194–206 (1997).
23. Mosbacher, J. *et al.* P2Y receptor subtypes differentially couple to inwardly rectifying potassium channels. *FEBS Lett.* **436**, 104–110 (1998).
24. von Heijne, G. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14**, 4683–4690 (1986).
25. O'Hara, P. J. *et al.* The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. *Neuron* **11**, 41–52 (1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank D. Ristig, A. Begrich, I. Meigel and S. Leonhard for technical assistance.

Correspondence and requests for materials should be addressed to B.B. (e-mail: bernhard.bettler@pharma.novartis.com). The rat GABA_B R2 cDNA sequence has been deposited at EMBL under the accession number AJ011318.

Moderate loss of function of cyclic-AMP-modulated KCNQ2/KCNQ3 K⁺ channels causes epilepsy

Björn C. Schroeder*, Christian Kubisch*, Valentin Stein & Thomas J. Jentsch

Zentrum für Molekulare Neurobiologie Hamburg (ZMNH), Universität Hamburg, Martinistrasse 85, D-20246 Hamburg, Germany

* These authors contributed equally to this work

Epilepsy affects more than 0.5% of the world's population and has a large genetic component¹. It is due to an electrical hyperexcitability in the central nervous system. Potassium channels are important regulators of electrical signalling, and benign familial neonatal convulsions (BFNC), an autosomal dominant epilepsy of infancy, is caused by mutations in the *KCNQ2* or the *KCNQ3* potassium channel genes^{2–4}. Here we show that *KCNQ2* and *KCNQ3* are distributed broadly in brain with expression patterns that largely overlap. Expression in *Xenopus* oocytes indicates the formation of heteromeric *KCNQ2/KCNQ3* potassium channels with currents that are at least tenfold larger than those of the respective homomeric channels. *KCNQ2/KCNQ3* currents can be increased by intracellular cyclic AMP, an effect that depends on an intact phosphorylation site in the *KCNQ2* amino terminus. *KCNQ2* and *KCNQ3* mutations identified in BFNC pedigrees compromised the function of the respective subunits, but exerted no dominant-negative effect on *KCNQ2/KCNQ3* heteromeric channels. We predict that a 25% loss of heteromeric *KCNQ2/KCNQ3*-channel function is sufficient to cause the electrical hyperexcitability in BFNC. Drugs raising intracellular cAMP may prove beneficial in this form of epilepsy.

We cloned the complete complementary DNA of the human *KCNQ3* K⁺ channel by homology to *KCNQ1* (also known as *KVLQT1* (ref. 5)) and determined its genomic structure. We identified two intronic CA nucleotide repeats which represent microsatellite markers linked previously to the long arm (q) of chromosome 8 at band 24, one of the known loci^{6,7} for BFNC. Indeed, a partial *KCNQ3* cDNA was cloned recently and a *KCNQ3* mutation was identified in a BFNC family³. The *KCNQ3* protein (Fig. 1) is 41% identical to *KCNQ2*, the other K⁺ channel mutated in BFNC, and 31% identical to *KCNQ1*, a K⁺ channel subunit mutated in the long QT syndrome⁵ (LQTS). It displays the typical structure of a K⁺ channel with six transmembrane domains and a pore-forming P-loop.

KCNQ3 shows similarities to *KCNQ2* (ref. 2) in being highly specific for brain (Fig. 2a). Northern blot analysis and *in situ* hybridization revealed that both genes are co-expressed in most brain regions, but there are some regional quantitative differences (Fig. 2b). Potassium channels are tetramers of identical or homologous α -subunits^{8–10}, raising the possibility that *KCNQ2* and *KCNQ3* form heteromeric channels. After co-expressing *KCNQ2* and *KCNQ3* that had been N-terminally tagged with haemagglutinin (HA) or Flag epitopes in *Xenopus* oocytes, we were able to co-

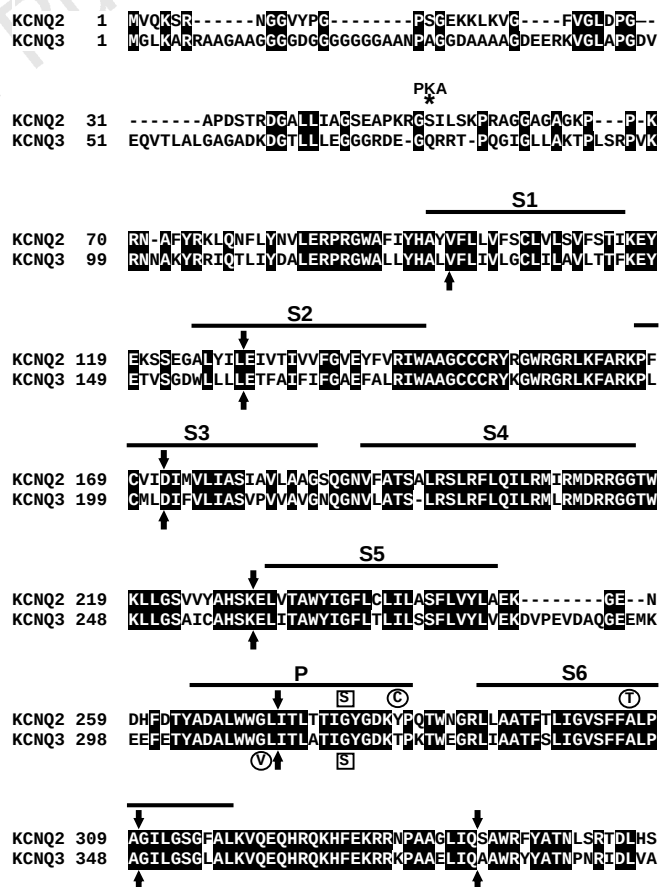


Figure 1 Partial alignment of *KCNQ2* and *KCNQ3* potassium channel subunits showing the newly identified *KCNQ3* N terminus, the cAMP-dependent phosphorylation site (*) in *KCNQ2*, and the point mutations introduced here. Identical residues are marked by black background. Predicted transmembrane spans S1–S6 and the P-loop (P) are indicated. Positions of introns in both genes are shown by arrows, but the *KCNQ2* exon-intron structure is incomplete². The Y284C and A306T (*KCNQ2*) and G310V (*KCNQ3*) mutations (circles) were found^{3,4} in BFNC pedigrees, whereas the G279S (*KCNQ2*) and G318S (*KCNQ3*) mutations (squares) were modelled on mutations in *KCNQ1* found¹⁴ in dominant LQTS patients.

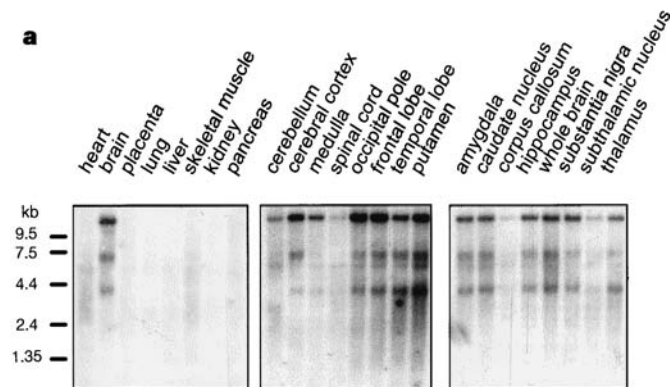
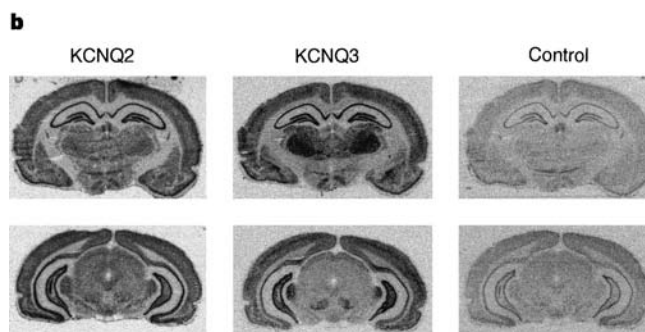


Figure 2 Expression patterns of KCNQ2 and KCNQ3. **a**, Northern blot analysis of KCNQ3 expression. Among the human tissues analysed, KCNQ3 mRNA was found only in brain (left panel). Within brain (centre and right panels), expression occurs in most regions, but transcript levels are low in the spinal cord and corpus callosum. **b**, *In situ* hybridization of rat brain slices with antisense probes directed



against KCNQ2 (left) and KCNQ3 (centre). Right panel shows control hybridization with a KCNQ2 sense probe. KCNQ2 and KCNQ3 expression overlaps in large areas of the brain, but KCNQ3 expression is higher in several nuclei, for example the amygdala and thalamic nuclei.

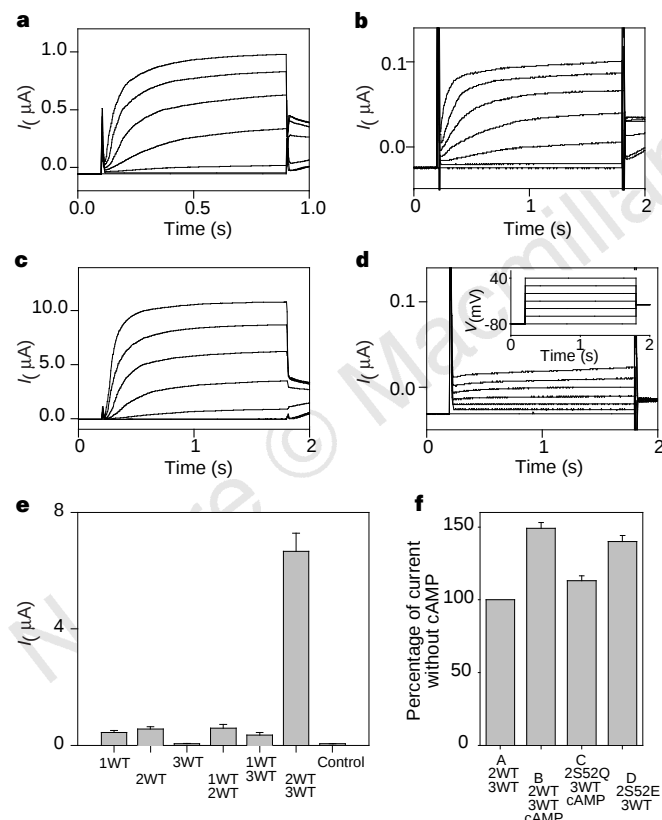


Figure 3 Functional expression of KCNQ channels in *Xenopus* oocytes. **a**, Voltage-clamp traces of KCNQ2. **b**, Traces of KCNQ3. **c**, 1:1 co-expression of KCNQ2 and KCNQ3 (note different scales). **d**, Control oocyte (clamp protocol shown in inset). **e**, Mean currents at +40 mV of different KCNQs (for example, 2WT means KCNQ2 wild type). Measurements are from a single batch of oocytes ($n > 8$, error bars indicate s.e.m.). **f**, Effect of raising intracellular cAMP (see Methods) on currents measured as in **e** (error bars indicate s.e.m.). Currents were normalized to those of the respective oocyte before applying cAMP (bars B, C), or to oocytes of the same batch injected with WT KCNQ2/KCNQ3 (bar D). 2S52Q and 2S52E indicate KCNQ2 mutants deleted for the PKA site or mimicking a phosphorylated PKA site, respectively.

precipitate KCNQ2 with KCNQ3 and vice versa (data not shown).

When expressed in *Xenopus* oocytes, KCNQ3 elicited currents (Fig. 3b) that were only barely above background (Fig. 3d), but resembled the larger depolarization-activated K^+ currents observed² with KCNQ2 (Fig. 3a). Co-injection of both cRNAs yielded currents (Fig. 3c) that were at least one order of magnitude larger than those observed with either subunit alone (Fig. 3e). Their voltage dependence and kinetics roughly resembled those of KCNQ1, KCNQ2 and KCNQ3, but seemed less inwardly rectifying than KCNQ2 currents. Currents observed after KCNQ1/KCNQ2 co-expression, however, did not differ significantly from a linear superposition of currents from the respective homomeric channels. KCNQ1 currents are changed markedly by co-expressing minK (IsK)^{11,12}, but we observed no pronounced effect when we expressed KCNQ2 and KCNQ3 together with this β -subunit at a concentration (0.2 ng per oocyte) sufficient to change KCNQ1 currents (data not shown). The relevance of such an effect would, however, be questionable because minK is not expressed in significant amounts in brain¹³.

Because the cytoplasmic N terminus of KCNQ2 contains a consensus site for cAMP-dependent phosphorylation (Fig. 1), we raised the intracellular cAMP concentration by applying cpt-cAMP (8-(4-chlorophenylthio)-cAMP), forskolin and IBMX (3-isobutyl-1-methylxanthine). This reversibly increased currents with wild-type (WT) KCNQ2/KCNQ3 by $49 \pm 16\%$ ($\pm$ s.d.) ($n = 17$) (Fig. 3f, bar B). This increase was largely inhibited by H-89, a blocker of protein kinase A (PKA) (data not shown). A mutation (S52Q) eliminating the potential N-terminal phosphorylation site of KCNQ2 reduced the stimulation by cAMP to $13 \pm 12\%$ ($\pm$ s.d.) ($n = 12$) (Fig. 3f, bar C), indicating that the current increase depends on the phosphorylation of this site. Whereas some oocytes injected with the mutant channel lacked a response to cAMP, in others a small response remained. This indicates that other processes (for example, phosphorylation of alternative sites or endo-exocytosis of channels) may participate in the current stimulation. When we introduced a mutation (S52E) in which the negatively charged glutamate mimics a phosphorylated serine residue, currents were larger by $\sim 40\%$ (25–65%, 128 oocytes from four different batches) (Fig. 3f, bar D). These experiments indicate that the activation of KCNQ2/KCNQ3 channels by cAMP is mediated by PKA. To show this more directly, we exposed excised patches from HEK293 cells expressing both channel subunits to the catalytic subunit of PKA in the presence of Mg-ATP. This led to a large, reversible enhancement of K^+ currents (Fig. 4).

We next investigated the effects of point mutations on the function of homo- and heteromeric channels to understand further

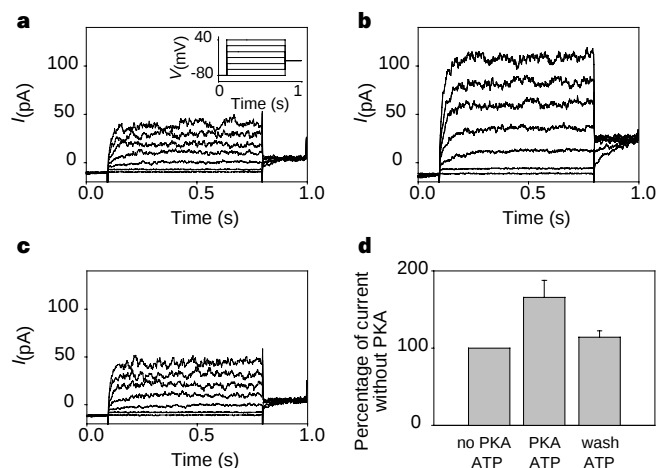


Figure 4 Effect of PKA on KCNQ2/KCNQ3 channels expressed in HEK293 cells. **a–c**, Currents from an inside-out patch in presence of 2 mM Mg-ATP before addition of PKA (**a**), 1 min after addition of PKA (**b**), and after wash-out (**c**). Clamp protocol is shown in the inset in **a**. **d**, Effect of PKA on currents measured as in **a–c** ($n = 6$, error bars indicate s.e.m.). PKA stimulated currents by $66 \pm 22\%$ (centre bar). This effect was largely reversible as currents returned to $14 \pm 8\%$ above the original level after wash-out (right bar). Currents were normalized to the respective current before applying PKA (left bar).

the pathomechanism of epilepsy. Y284C, a BFNC mutation⁴ in the pore region of KCNQ2, did not abolish channel function. Currents were reduced to ~30–60% (not shown), and a 1:1 co-expression with KCNQ3 again yielded much larger currents (Fig. 5a, bar D). To mimic the situation in a (heterozygous) patient with this dominant disease, we injected KCNQ2(Y284C), KCNQ2, and KCNQ3 at a ratio of 1:1:2 (Fig. 5a, bar B). Compared with wild-type KCNQ2/KCNQ3 heteromers, this combination reduced currents by only 20–30%. The only known BFNC mutation in KCNQ3, G310V, also affects a residue in the pore region³. We could not detect currents with KCNQ3(G310V), but in view of the low currents of wild-type KCNQ3 this does not prove a total loss of function. When co-injected 1:1 with KCNQ2, we observed large currents that were ~50% of wild-type heteromeric currents (Fig. 5b, bar L). A co-injection scheme mimicking the situation in a heterozygous patient (Fig. 5b, bar J) again resulted in currents reduced by only 20% compared with wild type. In addition to these pore mutations, we tested the effect of a KCNQ2 missense mutation (A306T)⁴ in transmembrane domain S6 and of an insertion of five base pairs (1592ins5bp)² in the carboxy terminus of KCNQ2 that leads to a truncated protein. Both mutations cause BFNC (refs 2, 4). Again, when co-expressed with wild-type KCNQ2 and KCNQ3 at a 1:1:2 ratio, the reduction in currents was in the range of 20–40% (Fig. 5c, bars R, S). On the basis of titration experiments (not shown), similar reductions in currents would be expected in patients with a deletion of KCNQ2 on one allele⁴. In all cases, we did not observe physiologically significant changes in ion selectivity or gating kinetics with mutant/wild-type heteromeric channels. This includes heteromeric channels with mutations in the P-loop, which retained the high K^+/Na^+ permeability ratio (>40) of wild-type channels. Thus, none of these BFNC mutations exerted a dominant-negative effect (defined by a current reduction of $>50\%$).

To contrast the behaviour of BFNC mutants with a dominant-negative effect, we constructed KCNQ2(G279S) and KCNQ3(G318S). These mutations were modelled on KCNQ1(G291S), a pore mutation that is found¹⁴ in the dominant Romano–Ward LQTS and which has a dominant-negative effect¹⁵. In contrast to the BFNC mutants studied above, these mutants exerted a dominant-

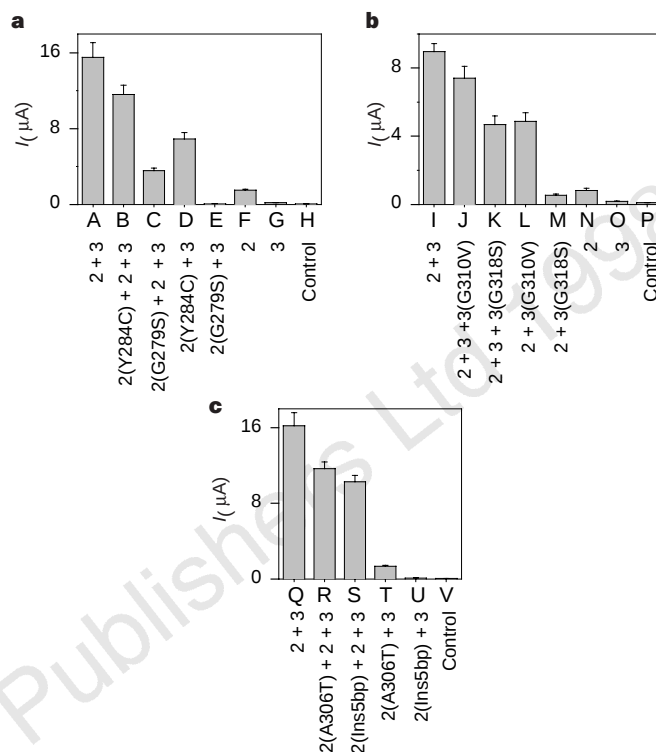


Figure 5 Functional analysis of KCNQ2 and KCNQ3 mutations. Effect of pore mutations in KCNQ2 (**a**) or KCNQ3 (**b**) and of non-pore mutations in KCNQ2 (**c**) on KCNQ2/KCNQ3 currents (same total amount of RNA). KCNQ2(Y284C), KCNQ2(A306T), KCNQ2(ins5bp) and KCNQ3(G310V) are BFNC mutations^{2–4}, whereas KCNQ2(G279S) and KCNQ3(G318S) mimic a dominant-negative KCNQ1 mutation^{14,15}. **a**, KCNQ2(G279S) exerts a strong dominant-negative effect with (bar C) or without (bar E) wild-type KCNQ2. The BFNC mutation KCNQ2(Y284C) lacks a dominant-negative effect (bars B, D), as does KCNQ3(G310V) (**b**, bars J, L). The artificial KCNQ3(G318S) mutation has a dominant-negative effect with (bars K, M) (smaller than the one of KCNQ2(G279S) in **a**). In **c**, a mutation in S6 (A306T)⁴ and a truncation (1592ins5bp)² reduce currents by $<40\%$ (bars R and S, respectively). The KCNQ2(ins5bp) mutant did not enhance KCNQ3 currents (bar U). Different batches of oocytes were used in **a**, **b** and **c**. Within each panel, a single batch was used. Bars, mean values from >11 oocytes; error bars, s.e.m.

negative effect on heteromeric channels both in the presence and in the absence of the respective wild-type subunit (Fig. 5a, b). The inhibitory effect of the KCNQ2 mutant was somewhat larger.

The overlapping expression pattern, the ability to associate physically, the large increase in currents upon co-expression, and the observation that mutations in KCNQ2 and KCNQ3 cause clinically identical diseases suggest that BFNC is due to a loss of KCNQ2/KCNQ3 heteromeric channels. Surprisingly, our experiments predict that this loss of function is quite small. This applies for missense mutations as well as for C-terminal deletions and gene deletions, both of which are rather common mutation types^{2,4} in BFNC. This is in contrast to the dominant LQTS, where mutations in KCNQ1 act in a dominant-negative manner^{13,15}. It is tempting to speculate that dominant-negative mutations in KCNQ2 or KCNQ3 may lead to more severe phenotypes. The slight loss of channel function predicted for BFNC patients implies that expression levels of KCNQ2/KCNQ3 channels are at a critical level during the early period of life in which this transient form of epilepsy occurs. The fact that the stimulation of KCNQ2/KCNQ3 currents by cAMP is within the range of the reduction of currents in BFNC suggests new ways in which this and possibly other forms of epilepsies may be treated pharmacologically. □

Methods

Cloning of KCNQ3 and generation of mutants. Expressed sequence tags (IMAGE Consortium cDNA clones¹⁶, accession numbers AA001392, AA019129 and H86050) homologous to KCNQ1 were extended at both ends by screening human brain cDNA libraries and by polymerase chain reaction (PCR) after reverse transcription of RNA (RT-PCR) techniques. The start ATG was assigned to the first ATG in frame following an upstream stop codon. The cDNA was inserted into the expression vector PTLN (ref. 17) which uses *Xenopus* β -globin untranslated sequences to boost expression in *Xenopus* oocytes. The genomic structure of KCNQ3 was determined by amplifying genomic sequences from human DNA using exonic primers and sequencing the products. The sequence of KCNQ3 exons and adjacent intronic stretches were deposited in the GenBank/EMBL database (accession numbers AF071478–AF071491). In introns 1 and 7, we discovered the microsatellite markers D8S558 and D8S1835 that map to chromosome 8q24. Mutagenesis was done using recombinant PCR and the *Pfu* DNA polymerase. Constructs were verified by sequencing using ABI 377 or 310 automated sequencers.

Analysis of tissue distribution. For northern blot analysis, we used human RNA blots (Clontech) and a radiolabelled 526-base-pair *NotI/HindIII* fragment from KCNQ3 as a probe. For *in situ* hybridization, 1.5-kilobase antisense and control sense RNA probes labelled with ³⁵S were transcribed by T3 or T7 RNA polymerases from KCNQ2 and KCNQ3 partial clones in pBluescript using the Maxiscript kit (Ambion). Template DNA and unincorporated nucleotides were removed. Probe sizes were reduced by controlled alkaline treatment. *In situ* hybridization was done as described¹⁸, with the following modifications: after washing in PBS, sections were acetylated without ultraviolet crosslinking, and were exposed to Kodak Biomax X-ray film.

Expression in *Xenopus* oocytes and HEK cells and electrophysiology. After linearization of the constructs, capped cRNA was synthesized using SP6 RNA polymerase and the mMessage mMachine kit (Ambion). For KCNQ1, we used isoform 0 (ref. 15). *Xenopus* oocytes were prepared, injected (generally with 5 ng cRNA), and handled as described¹⁹. After 1–2 days, currents were examined at room temperature in saline containing (in mM) 96 NaCl, 2 KCl, 3 MgCl₂, 0.2 CaCl₂, and 5 HEPES (pH 7.4) using two-electrode voltage clamping with a TurboTec amplifier (NPI). To raise intracellular cAMP, we applied 200 μ M cpt-cAMP, 10 μ M forskolin and 1 mM IBMX for 15 min. For kinase inhibition, oocytes were exposed for >6 h to 100 μ M H-89 (Calbiochem). HEK 293 cells were transiently transfected with KCNQ2 and KCNQ3 cDNAs subcloned into pCDNA3 (Invitrogen) using Lipofectamin (Gibco). The inside-out configuration of the patch-clamp technique was used with borosilicate glass pipettes (5 M Ω resistance) and an Axopatch 200-A amplifier (Axon). Currents were filtered at 250 Hz, digitally sampled at 500 Hz, and analysed with pClamp (Axon) and Sigma Plot (Jandel Scientific) software. The bath solution contained, in mM: 140 K, 140 aspartic acid, 2 MgCl₂, 5 HEPES, 2 EGTA (pH 7.3 adjusted with KOH); pipette solution: 140 Na, 5 K, 145 aspartic acid, 2 CaCl₂, 2 MgCl₂, 5 HEPES (pH 7.3 adjusted with NaOH). Excised patches were superfused with bath solution containing 2 mM Mg-ATP or 2 mM Mg-ATP and 40 U m⁻¹ PKA catalytic subunit from bovine heart (Calbiochem) where noted.

Received 21 September; accepted 3 November 1998.

- Baraitser, M. In *The Genetics of Neurological Disorders* 2nd edn (Oxf. Monogr. Med. Genet., No. 18) 96–113 (Oxford Univ. Press, New York, 1990).
- Biervet, C. et al. A potassium channel mutation in neonatal human epilepsy. *Science* **279**, 403–406 (1998).
- Charlier, C. et al. A pore mutation in a novel KQT-like potassium channel in an idiopathic epilepsy family. *Nature Genet.* **18**, 53–55 (1998).
- Singh, N. A. et al. A novel potassium channel gene, *KCNQ2*, is mutated in an inherited epilepsy of newborns. *Nature Genet.* **18**, 25–29 (1998).
- Wang, Q. et al. Positional cloning of a novel potassium channel gene: *KVLQT1* mutations cause cardiac arrhythmias. *Nature Genet.* **12**, 17–23 (1996).
- Leppert, M. et al. Benign familial neonatal convulsions linked to genetic markers on chromosome 20. *Nature* **337**, 647–648 (1989).
- Lewis, T. B., Leach, R. J., Ward, K., O'Connell, P. & Ryan, S. G. Genetic heterogeneity in benign familial neonatal convulsions: identification of a new locus on chromosome 8q. *Am. J. Hum. Genet.* **53**, 670–675 (1993).
- Christie, M. J., North, R. A., Osborne, P. B., Douglass, J. & Adelman, J. P. Heteropolymeric potassium channels expressed in *Xenopus* oocytes from cloned subunits. *Neuron* **2**, 405–411 (1990).
- Isacoff, E. Y., Jan, Y. N. & Jan, L. Y. Evidence for the formation of heteromultimeric potassium channels in *Xenopus* oocytes. *Nature* **345**, 530–534 (1990).
- Ruppersberg, J. P. et al. Heteromultimeric channels formed by rat brain potassium-channel proteins. *Nature* **345**, 535–537 (1990).
- Barhanin, J. et al. KvLQT1 and Isk (minK) proteins associate to form the *I_{Ks}* cardiac potassium current. *Nature* **384**, 78–80 (1996).
- Sanguinetti, M. C. et al. Coassembly of KvLQT1 and minK (IsK) proteins to form cardiac *I_{Ks}* potassium channel. *Nature* **384**, 80–83 (1996).

- Chouabe, C. et al. Properties of KvLQT1 K⁺ channel mutations in Romano-Ward and Jervell and Lange-Nielsen inherited cardiac arrhythmias. *EMBO J.* **16**, 5472–5479 (1997).
- Russell, M. W., Macdonald, D., Collins, F. S. & Brody, L. C. KVLQT1 mutations in three families with familial or sporadic long QT syndrome. *Hum. Mol. Genet.* **5**, 1319–1324 (1996).
- Wollnik, B. et al. Pathophysiological mechanisms of dominant and recessive KVLQT1 K⁺ channel mutations found in inherited cardiac arrhythmias. *Hum. Mol. Genet.* **6**, 1943–1949 (1997).
- Lennon, G., Aufray, C., Polymeropoulos, M. & Soares, M. B. The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* **33**, 151–152 (1996).
- Lorenz, C., Pusch, M. & Jentsch, T. J. Heteromultimeric CLC chloride channels with novel properties. *Proc. Natl Acad. Sci. USA* **93**, 13362–13366 (1996).
- Hartmann, D., Fehr, S., Meyerhof, W. & Richter, D. Distribution of somatostatin receptor subunit 1 mRNA in the developing cerebral hemispheres of the rat. *Dev. Neurosci.* **17**, 246–255 (1995).
- Jordt, S.-E. & Jentsch, T. J. Molecular dissection of gating in the CLC-2 chloride channel. *EMBO J.* **16**, 1582–1592 (1997).

Acknowledgements. We thank S. Fehr for the preparation of rat brain slices and the *in situ* hybridization, and T. Friedrich for comments on the manuscript. This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie to T.J.J.

Correspondence and requests for materials should be addressed to T.J.J. (e-mail: jentsch@plexus.uke.uni-hamburg.de).

Changes in thymic function with age and during the treatment of HIV infection

Daniel C. Douek*, Richard D. McFarland†, Philip H. Keiser*, Earl A. Gage*, Janice M. Massey‡, Barton F. Haynes‡, Michael A. Polis§, Ashley T. Haase||, Mark B. Feinberg¶, John L. Sullivan#, Beth D. Jamieson**, Jerome A. Zack***, Louis J. Picker† & Richard A. Koup*

Departments of * Medicine and † Pathology, The University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA

‡ Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710, USA

§ National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

|| Department of Microbiology, University of Minnesota Medical School, Minneapolis, Minnesota 55455, USA

¶ Department of Medicine, Emory University School of Medicine, Atlanta, Georgia 3022, USA

Departments of Pediatrics and Molecular Medicine, The University of Massachusetts Medical Center, Worcester, Massachusetts 01605, USA

** Division of Hematology-Oncology, Department of Medicine, and

*** Department of Microbiology and Molecular Genetics, UCLA School of Medicine and UCLA Institute, Los Angeles, California 90095, USA

The thymus represents the major site of the production and generation of T cells expressing $\alpha\beta$ -type T-cell antigen receptors¹. Age-related involution² may affect the ability of the thymus to reconstitute T cells expressing CD4 cell-surface antigens that are lost during HIV infection³; this effect has been seen after chemotherapy and bone-marrow transplantation^{4,5}. Adult HIV-infected patients treated with highly active antiretroviral therapy (HAART) show a progressive increase in their number of naive CD4-positive T cells^{6,7}. These cells could arise through expansion of existing naive T cells in the periphery⁸ or through thymic production of new naive T cells^{9,10}. Here we quantify thymic output by measuring the excisional DNA products of TCR-gene rearrangement. We find that, although thymic function declines with age, substantial output is maintained into late adulthood. HIV infection leads to a decrease in thymic function that can be measured in the peripheral blood and lymphoid tissues. In adults treated with HAART, there is a rapid and sustained increase in thymic output in most subjects. These results indicate that the adult thymus can contribute to immune reconstitution following HAART.

In humans, there is no known way to distinguish phenotypically between cells that have recently emigrated from the thymus and long-

lived naive cells in the periphery, and therefore to quantify thymic output. One proposed marker for recent thymic emigrants (RTEs) is the episomal DNA circles that are generated during excisional rearrangement of TCR genes¹¹. These products, termed TCR-rearrangement excision circles (TRECs), are stable¹², are not duplicated during mitosis, and are therefore diluted out with each cellular division¹³. In T cells expressing TCR- $\alpha\beta$ ($\alpha\beta$ T cells), rearrangement of both *TCRA* and *TCRB* genes produces TRECs¹⁴. Because of the enormous diversity of *TCRAVJ* and *TCRBVDJ* recombination events, and thus of the TRECs produced, no single TREC can be used as a marker to assess overall thymic function. However, a common requirement for productive rearrangement of the *TCRA* locus is deletion of the *TCRD* locus which it encompasses (Fig. 1a). The two rearrangement events that occur during this process are identical in ~70% of $\alpha\beta$ T cells¹⁵, the first producing a signal-joint TREC and the second a coding-joint TREC (Fig. 1a). Thus the TRECs generated are common to most $\alpha\beta$ T cells and can be used to detect and quantify thymic output of $\alpha\beta$ T cells in clinical samples.

To determine the cell specificity of signal-joint and coding-joint TRECs, we purified phenotypically distinct lymphocyte populations by fluorescence-activated cell sorting (FACS) from adult peripheral blood mononuclear cells (PBMCs) and used the polymerase chain reaction (PCR) to detect signal- and coding-joint sequences. We detected these TRECs in phenotypically naive T cells, but not in memory T cells, B cells or TCR- $\gamma\delta$ T cells (Fig. 1b). These results confirm that signal-joint and coding-joint sequences are unique to naive $\alpha\beta$ T cells. In addition, the fact that memory T cells lack these specific sequences indicates that the TRECs may be diluted during the switch from naive to memory phenotype, probably as a result of the cellular proliferation that accompanies this process¹⁶, and that neither sequence is retained within the germ line after TCR-gene recombination, consistent with a lack of allelic exclusion during *TCRA* locus recombination¹⁷. We created a DNA standard that was identical to the region amplified by the signal-joint primers, but containing a 60-base pair (bp) internal deletion. We used this internal standard to quantify signal-joint TRECs in thymocytes and PBMCs by quantitative competitive (QC)-PCR¹⁸ (Fig. 1c). QC-PCR was further used to confirm that signal-joint TRECs do not replicate during mitosis and are diluted during *in vitro* cell division (Fig. 1d). For every $\log_{10}$ increase in CD4⁺ or CD8⁺ T-cell number after stimulation with anti-CD3 and anti-CD28 monoclonal antibody, there was an equivalent decrease in signal-joint TRECs per microgram of cellular DNA. Therefore, as cell number increased, the total number of signal-joint TRECs remained unchanged, leading to a decrease in signal-joint TRECs on a per cell basis.

To determine the relationship between age and TREC quantity, we isolated CD4⁺ and CD8⁺ T cells from PBMCs of subjects from birth (cord blood) to 73 years of age, and quantified signal- and coding-joint TRECs by QC-PCR. The numbers of both coding- and signal-joint TRECs decline exponentially in CD4⁺ and CD8⁺ T cells with the age of the donor (Fig. 2a). Over the life of a person, there is a 1–1.5 $\log_{10}$ drop in the number of TRECs whereas the corresponding drop in the number of cells with a naive phenotype is only roughly fourfold (Fig. 2b). Moreover, from 20–60 years of age the percentage of naive cells remains fairly constant¹⁹ whereas the number of TRECs continues to decline exponentially, indicating that the presence of TRECs is not simply a surrogate for cells with a naive phenotype. Within cord-blood CD4⁺ and CD8⁺ T cells (100% naive phenotype), there were ~200,000 coding-joint TRECs per microgram DNA (~150,000 cells). Assuming that *TCRA* recombination occurs in both alleles¹⁷, this result would indicate that 200,000 of the potential 300,000 (67%) *TCRA/D* rearrangements produced signal- and coding-joint TRECs. This is close to what has been estimated previously by other techniques¹⁵, and indicates that this assay detects most RTEs. There were routinely 8- to 16-fold more coding-joint TRECs than signal-joint TRECs in any sample, indicating that approximately 3–4 cell divisions occur between the

TCRD and the *TCRA* locus rearrangements, which produce signal-joint and coding-joint TRECs, respectively. During these cell divisions, the signal-joint TRECs would not replicate and would be diluted while the coding-joint sequence would remain part of the genomic DNA. Because signal- and coding-joint TRECs are affected similarly in mature T cells, we quantified only signal-joint TRECs in

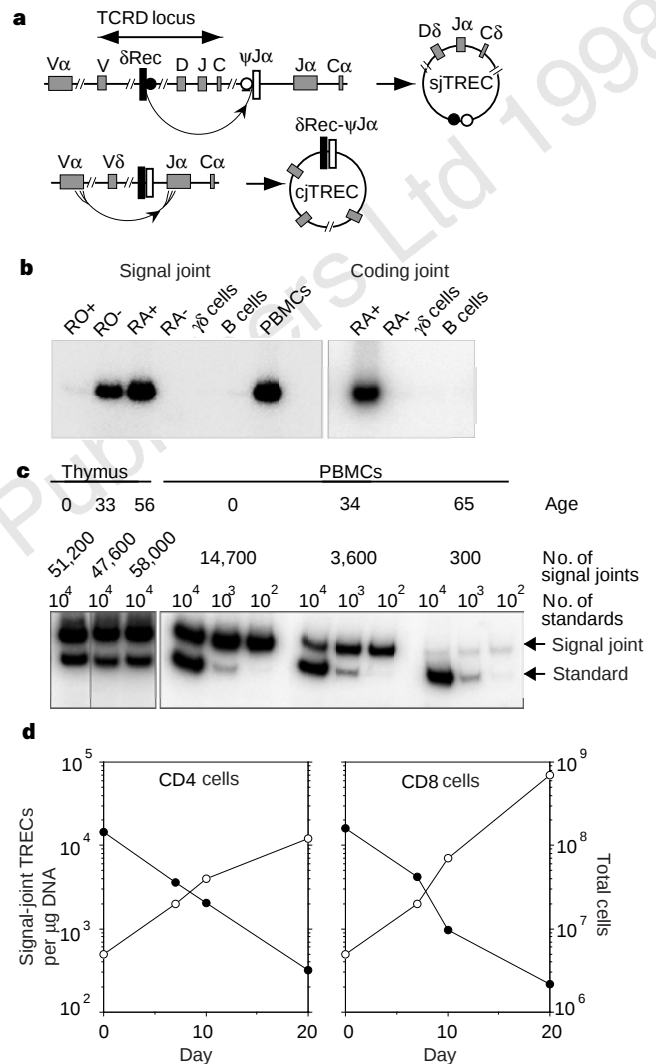


Figure 1 Generation of signal-joint and coding-joint TRECs and their detection in lymphocyte populations. **a**, Top, a simplified representation of the *TCRD* locus flanked by portions of the *TCRA* locus (labelled as V α , J α and C α). End-to-end ligation of the recombination signal sequences flanking the δ -rec locus and the ψ -J α locus removes most of the *TCRD* gene region, forming a single TREC containing a unique signal joint (sj) sequence. Bottom, the recombined δ -rec to ψ -J α junction (the coding joint, cj) is retained in the germline DNA until *TCRAV* to *TCRAJ* recombination occurs; the coding joint then becomes part of a second TREC. As the δ -rec to ψ -J α recombination cannot produce a functional TCR, signal- and coding-joint sequences can only occur in TRECs in mature T cells. **b**, Presence of signal- and coding-joint TRECs in T-cell subpopulations sorted by FACS and in unsorted PBMCs. CD45RA⁺ (RA⁺) and CD45RO⁺ (RO⁺) cells are naive; CD45RA⁺ (RA⁺) and CD45RO⁺ (RO⁺) cells are memory CD4⁺ T cells. **c**, QC-PCR for signal-joint TRECs. The number of signal-joint TRECs in MACS-purified CD4⁺ cells (PBMCs) of two representative adult individuals (aged 34 and 65 years) and one cord-blood sample (aged 0), and in unsorted thymocytes from a fetus and adults aged 33 and 56 years. The age of each subject, the number of signal joints per microgram DNA, and the number of molecules of internal standard added to each PCR reaction are shown. **d**, Signal-joint TRECs (filled circles) decline with cellular proliferation (open circles) during successive days after *in vitro* stimulation of sorted CD4⁺ and CD8⁺ cord-blood T cells.

subsequent experiments.

The continued presence of TREC-containing cells within the peripheral blood of elderly subjects could reflect these cells' long-

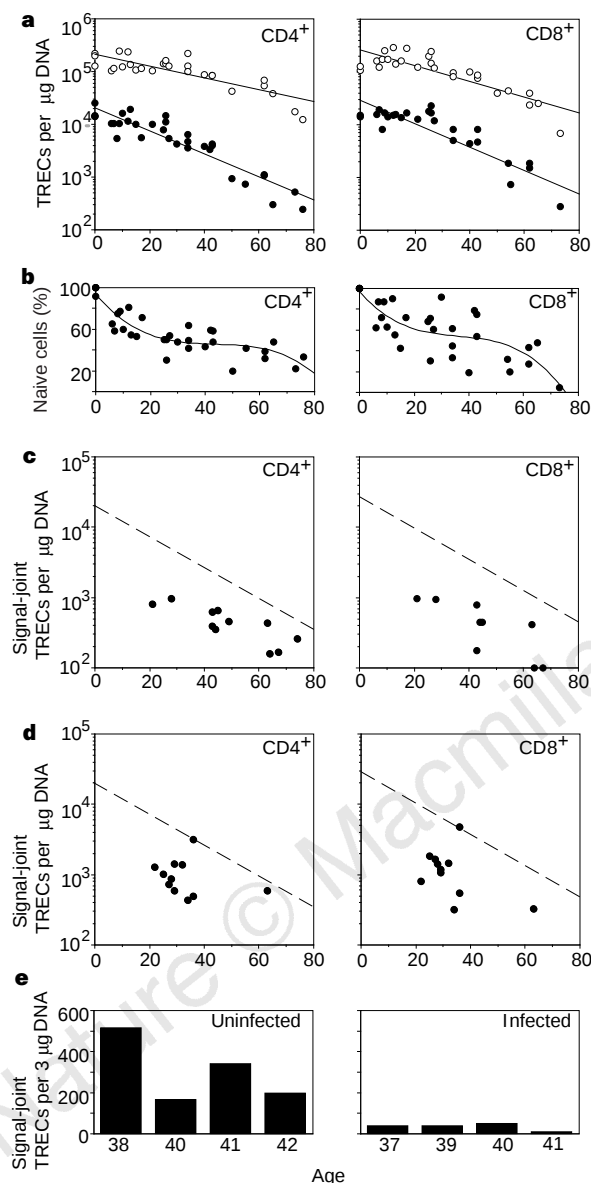


Figure 2 TREC levels in healthy, athymic and HIV-infected individuals. **a**, Graphs show the number of coding-joint of TRECs (open circles) and signal-joint TRECs (filled circles) in MACS-sorted CD4^+ (left) and CD8^+ (right) cells from healthy volunteers of different ages. Best-fit exponential curves are shown ($P < 0.001$ for all graphs). **b**, Percentages of naive cells in CD4^+ and CD8^+ populations in the healthy volunteers used for **a**. Best-fit cubic curves are shown (CD4^+ : $R^2 = 0.8$, $P < 0.001$; CD8^+ : $R^2 = 0.6$, $P = 0.06$, not significant). **c**, Signal-joint TREC levels in MACS-sorted CD4^+ and CD8^+ cells from thymectomized individuals (filled circles). The linear regression curve for the healthy individuals in **a** is shown (dashed line). There was a significant difference between the TREC levels in CD4^+ cells of healthy and thymectomized individuals below age 78 ($P < 0.001$), and between TREC levels in CD8^+ cells of healthy and thymectomized individuals at all ages ($P < 0.001$). **d**, Signal-joint TRECs in MACS-sorted CD4^+ and CD8^+ cells from HIV-infected subjects (filled circles) before HAART. The linear regression curve for the healthy individuals in **a** is shown (dashed line). There was a significant difference between the TREC levels in CD4^+ cells of healthy and HIV-infected individuals below age 50 ($P < 0.001$), and between TREC levels in CD8^+ cells of healthy and HIV-infected individuals of all ages ($P < 0.001$). **e**, The number of signal-joint TRECs per $3 \mu\text{g}$ of lymph node DNA from HIV-infected and uninfected individuals of different ages.

evity but might also reflect the sustained output of RTEs from the thymus. To address this possibility we quantified signal-joint TRECs in T cells purified from PBMCs of subjects who had undergone total thymectomy 3–39 years previously, as treatment for myasthenia gravis or during cardiac surgery. Figure 2c shows that the amount of signal-joint TRECs in CD4^+ and CD8^+ cells in athymic subjects was roughly an order of magnitude lower than in euthymic patients and remained consistently lower, although in CD4^+ cells the difference became progressively less with increasing age. This indicates that continued thymic output may occur, maintaining a persistent level of TREC-containing RTEs well into adulthood. We also found high levels of TRECs within the thymocytes of older individuals, confirming that their thymi are still capable of producing thymocytes that actively rearrange TCR genes (Fig. 1c).

The observations that TREC numbers decline with cellular proliferation, with increasing age and after thymectomy indicate that TREC levels in peripheral blood T cells are a measure both of thymic output, which could increase TREC levels, and of peripheral expansion, which would decrease TREC levels. We investigated the effect of HIV infection on these parameters by quantifying TRECs in the peripheral blood cells of HIV-infected subjects, relatively early in the progression of their disease, both before and during HAART. TREC levels were significantly lower in both CD4^+ and CD8^+ cells after HIV infection, compared with age-matched healthy controls (Fig. 2d). TREC numbers were similarly reduced when calculated per naive T cell (data not shown). Levels of TRECs in lymph nodes from four untreated HIV-infected subjects with established disease were also lower than in lymph nodes from age-matched healthy controls (Fig. 2e). The decrease in TREC numbers in both CD4^+ and CD8^+ T-cell populations could be due to peripheral events, such as proliferation among naive cells leading to a dilution of TRECs, or preferential infection and death of RTEs. However, peripheral expansion of naive cells occurs slowly, if at all²⁰, and it is unlikely that RTEs are preferentially susceptible to HIV infection^{21,22}. A more likely explanation is that the decrease in TREC amounts is thymic in origin, and results from HIV-induced inhibition of thymic function, inhibition of thymocyte precursors^{23,24}, or infection and death of thymocytes that express CD4, as has been shown in severe combined immunodeficient (SCID)-hu mice^{25,26}. We think, therefore, that the decrease in TRECs in naive cells in HIV-infected individuals reflects decreased thymic production of new naive T cells.

To determine whether RTEs could contribute to the reconstitution of the CD4^+ T-cell population in response to HAART⁶, we quantified TRECs in peripheral blood CD4^+ cells isolated from the above cohort of HIV-infected subjects at time points up to 9 months after the initiation of HAART. Figure 3 shows the changes in viral load, naive CD4^+ cell count and naive CD4^+ cell TRECs for each subject over time. The TREC levels are presented in this way so that subtle changes in RTEs within the naive CD4^+ cell population itself can be assessed independently of fluctuations in the relative percentages of naive and memory cells, thus providing a more accurate measure of thymic function. In all but one of the subjects (J-63) there was a rise in the number of TRECs in CD4^+ naive cells after initiation of HAART and suppression of viral load, even when the number of naive CD4^+ cells did not increase. In some subjects the initial rise in TREC amounts was rapid, occurring within 4–16 weeks after initiation of HAART. Two subjects, A-36 and B-28, showed large, sustained rises in TREC numbers to normal levels. In the other subjects who showed responses to HAART, TREC levels remained raised but were below normal levels at the final time point measured. In subjects C-25 and F-36, viral breakthrough was associated with a rapid decline in TREC levels. C-25's TREC level remained low until he was switched to a different HAART drug regimen, at which point his TREC levels increased rapidly and his viral load fell. Subject F-36 failed to show a response in TREC levels during his first course of HAART, which was stopped for a period of 4 weeks. After HAART was restarted, F-36's TREC numbers

increased as the viral load fell; this was followed by an episode of viral breakthrough, also associated with a transient decrease in TREC levels.

The variability in the magnitude and kinetics of the rise in TREC levels among the subjects may reflect the normal variability in thymic size⁹ and baseline TREC levels among individuals, both infected and uninfected. All but one of the subjects (J-63) were aged between 22 and 36 years, making it difficult to correlate TREC response with age. However, it is notable that J-63, who showed no rise in TREC amounts over the period studied, is considerably older, at 63 years.

Recent data indicate that the initial increase in CD4⁺ cells after HAART may result from release of sequestered cells from lymph nodes²⁷. Although this may indeed occur, the fact that we fail to find a reservoir of TREC-containing T cells in the lymph nodes of HIV-infected subjects indicates that the rise in TREC levels seen during HAART may be due not to the release of naive CD4⁺ cells from lymph nodes but instead to RTE production by the thymus. Indeed, other studies have shown that, later in the course of HAART, proliferative response of T cells to antigens recover and TCR repertoires become normalized²⁸, and that renewed thymopoiesis

occurs in HIV-infected human thymic implants after antiviral therapy²⁹. It is possible that some pre-T cells might undergo TCR-gene rearrangement at an extrathymic site, such as lymph node³⁰. It is difficult to exclude this possibility experimentally, but we have measured TREC levels in the few peripheral blood CD4⁺ and CD8⁺ cells of an individual with congenital thymic aplasia (DiGeorge syndrome). No TRECs were detectable in these cells (data not shown); thus it is unlikely that an extrathymic site can contribute significantly to the generation of TRECs.

Our results indicate that thymic function in adults may be suppressed in HIV-infected individuals and can be improved by the reduction of viral load. After early life, thymopoiesis becomes progressively less efficient and so thymus-mediated reconstitution of naive cells and of the broad TCR repertoire of these cells is likely to be a slow process in adults. Therapies that directly improve thymic function in adults may increase the rate of immune reconstitution after HAART. Whereas analysis of cell-surface phenotypes shows only late increases in the total number of naive T cells during HAART^{6,7}, changes in TREC levels occur much earlier in treatment and are indicative of the thymic potential for reconstitution of naive T cells. □

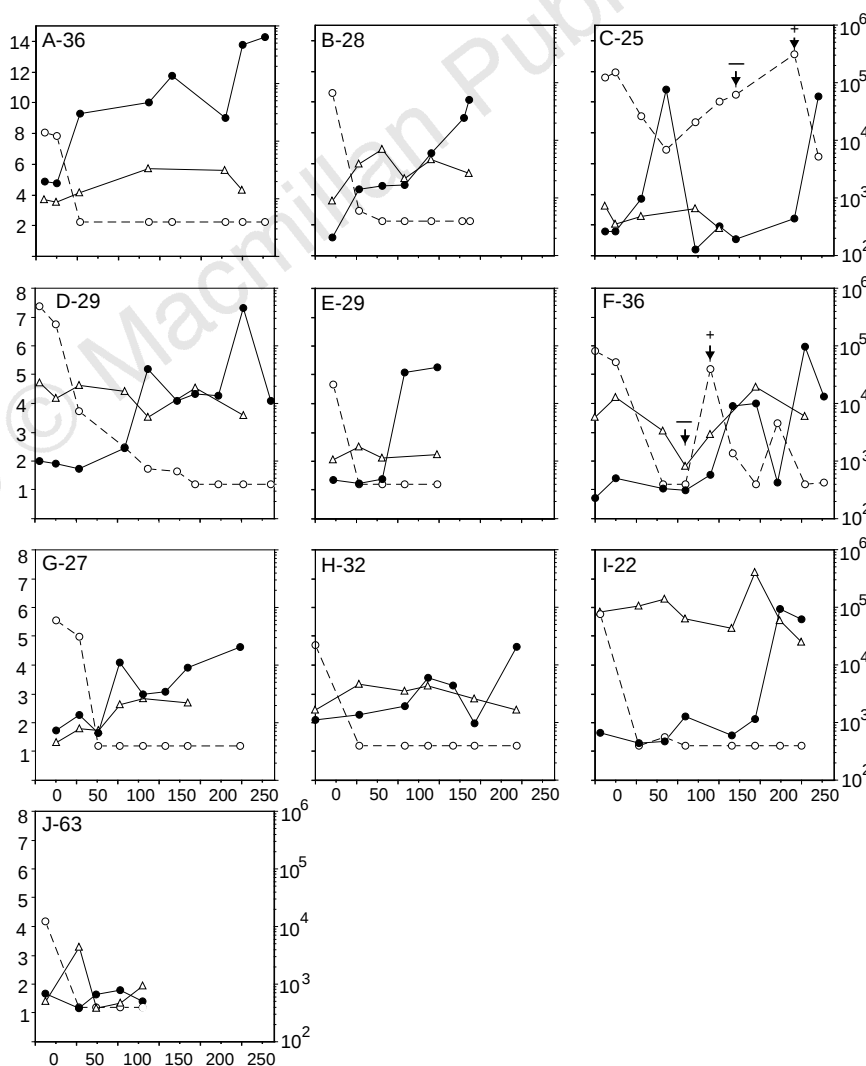


Figure 3 Changes in signal-joint TREC levels after HAART. The results plotted are: number of signal-joint TRECs $\times 10^3$ per microgram naive CD4⁺ cell DNA (filled circles, left y-axis); number of naive CD4⁺ T cells $\times 10^2$ per microlitre blood (triangles, left y-axis); number of HIV RNA copies per millilitre plasma (open circles, dashed line, right y-axis). The days from initiation of HAART (day 0) are

shown on the x-axis. Subjects are given a single-letter identification code, followed by their age, in the top left corner of each graph. Subjects C-25 and F-36 stopped therapy (–) and restarted therapy (+) at the times indicated with arrows. Note that the left y-axis scale for the top three graphs is different from the other scales, to accommodate the larger increases in signal-joint TREC levels.

Methods

FACS analysis. PBMCs from a 42-year-old subject were labelled with biotin- or fluorochrome-conjugated antibodies (Becton Dickinson and Caltag) against CD4, CD3, CD45RA, CD45RO, CD20 or TCR- $\gamma\delta$. Using a FACSort Plus (Becton Dickinson), we sorted these cells into populations of naive TCR- $\alpha\beta$ T cells (CD4⁺CD3⁺TCR- $\gamma\delta$ -CD45RA^{hi} and CD4⁺CD3⁺TCR- $\gamma\delta$ CD45RO⁻), memory T cells (CD4⁺CD3⁺TCR- $\gamma\delta$ -CD45RA⁻ and CD4⁺CD3⁺TCR- $\gamma\delta$ -CD45RO^{hi}), B cells (CD3⁺CD20⁺) and TCR- $\gamma\delta$ T cells, to a purity of >99%.

Magnetic cell sorting. PBMCs were separated into CD4⁺ and CD8⁺ populations using MACS CD4 and CD8 magnetic microbeads and positive selection columns (Miltenyi Biotech) according to the manufacturer's instructions. Purity of sorted populations was routinely >80% as determined by flow cytometry (data not shown).

PCR. Primer sequences are located ~180 bp on either side of the recombination joints. Primers for signal-joint TRECs, based on previously published primers¹⁵, were: 5'-aagaggcagccctcctcaaggcaaa-3' and 5'-aggctgat ctgtctgacattgtctcgcg-3'. Primers for coding-joint TRECs were: 5'-cctgtttgtaagg cacattagaatctctcactg-3' and 5'-ctaataaagatcctcaagggtcgagactgtc-3'. DNA was extracted from cells using Trizol (Gibco) and the DNA concentration was determined by spectrophotometry. The quality of some DNA samples was assessed by PCR with primers specific for a *TCRBC* region (data not shown). PCR conditions were: 95 °C for 5 min, followed by 90 °C, 60 °C and 72 °C, each for 30 s, for 30 or 35 cycles as indicated. Each PCR reaction contained 1 U platinum *Taq* polymerase (Gibco), 1.8 mM MgCl₂, 0.2 mM dNTPs, 12.5 μ M each primer and 1.7 nmol (5 μ Ci) ³²P-labelled dCTP in 50 μ l platinum *Taq* buffer.

QC-PCR. To each PCR reaction we added 10⁴, 10³ or 10² molecules of signal-joint internal standard, or 10⁵, 10⁴ or 10³ molecules of coding-joint internal standard. Internal standards had the same primer-binding sequences as the template for the target TREC but were genetically modified to be 60 bp shorter. We carried out exhaustive studies to determine the precise conditions under which PCR amplification was in the linear range of the assay. This occurred when 0.1–1 μ g of DNA extracted from purified T cells was amplified for between 30 and 35 cycles. PCR products were separated on non-denaturing 6% polyacrylamide gels. Bands were imaged and analysed using a Cyclone phosphorimager and Optiquant software (Packard). Because ³²P-labelled dCTP is incorporated in the reaction and there are fewer GC nucleotides in the internal standards (as determined by sequencing), the intensity of the standard band was multiplied by a correction factor of 1.156 for the signal-joint standard and 1.44 for the coding-joint standard. To quantify the number of TRECs in a PCR reaction, the target-TREC band intensity was divided by the corrected standard band intensity and multiplied by the number of standard molecules in the reaction. A constant number of TRECs could be calculated for each sample, irrespective of the number of standard molecules added (in the range 10⁴–10² standard molecules for signal joints and 10⁵–10³ for coding joints). We used 30-cycle QC-PCR unless 10² molecules of standard were added, when cycle number was increased to 35. Reproducibility of the assay was such that the TREC number calculated for the same sample never varied by more than 10% between different experiments.

Thymocytes. Single-cell suspensions were prepared from a fetal thymus at 18 weeks of gestation (Advanced Bioscience Resources) and thymus biopsies were taken from adults aged 33 and 56 years. Single-cell suspensions were washed once in PBS and lysed in urea lysis buffer as described²⁶.

In vitro stimulation of naive cells. PBMCs from a cord blood sample (in which almost all T cells are of naive phenotype) were separated into CD4⁺ and CD8⁺ populations using MACS microbeads. Cells were stimulated with 0.1 μ g ml⁻¹ anti-CD3 monoclonal antibody (clone 12F6) and 1 μ g ml⁻¹ anti-CD28 monoclonal antibody (Coulter) and were cultured in RPMI 1640 medium supplemented with 15% fetal bovine serum (Gibco) and 50 U ml⁻¹ recombinant interleukin-2. Cells were restimulated in the same way after 10 days. Total cell number was counted and an aliquot of each culture was reserved for DNA extraction on days 0, 7, 10 and 20 after the initial stimulation. The DNA was used to measure directly the level of signal-joint TRECs per microgram DNA.

HIV-infected subjects. Subjects were enrolled to receive HAART if they showed a positive antibody titre for HIV and had no history of receiving antiretroviral therapy. All subjects gave informed consent for the study.

Subjects were treated with several antiretroviral agents, including combinations of reverse-transcriptase inhibitors and/or protease inhibitors. Blood samples were taken before and during HAART at ~4-week intervals. Total CD4 count was measured and percentages of naive CD4⁺ T cells (CD45RO^{lo}, CD95^{lo}, CD27^{hi}) were assessed by flow cytometry (fluorochrome-conjugated antibodies were from Becton Dickinson). Plasma HIV load was measured (Amplicor, Roche); the limit of detection was 400 RNA copies per millilitre plasma. PBMCs were separated into CD4⁺ and CD8⁺ populations using MACS microbeads.

Lymph-node samples. DNA was extracted from formaldehyde-fixed, paraffin-embedded lymph nodes taken from untreated HIV-infected subjects with established disease and from routine biopsies (reported as histologically normal) of uninfected age-matched controls, using the QIAamp Tissue Kit (Qiagen). HIV-infected subjects were studied as part of a pathogenesis protocol at the NIH. All subjects gave informed consent before biopsy. QC-PCR was done with 3 μ g extracted DNA per sample, as T cells could not be purified from other lymphocytes. TREC levels in formaldehyde-fixed, paraffin-embedded tissues were at least fivefold lower than in fresh tissue (as determined with a thymocyte control; data not shown), as DNA extraction from fixed, paraffin-embedded tissue is much less efficient.

Statistical analysis. TREC levels were transformed to a log₁₀ scale; levels in normal individuals were compared with levels in individuals from either the thymectomized or the HIV-infected groups (adjusted for age), using analysis of covariance. When the assumption of homogeneity of slopes was not met, the Pothoff extension to the Johnson–Neyman technique was used to estimate the region of significance, in age, for which the two groups were statistically significant. Statistical analysis and curve fitting were done using SAS 6.12 software.

Received 18 September; accepted 9 November 1998.

- Pickler, L. & Siegelman, M. In *Fundamental Immunology* (ed. Paul, W. E.) 145–197 (Raven, New York, 1993).
- Steinmann, G. Changes in the human thymus during aging. *Curr. Top. Pathol.* **75**, 43–80 (1986).
- Pantaleo, G., Graziosi, C. & Fauci, A. The immunopathogenesis of human immunodeficiency virus infection. *New Engl. J. Med.* **328**, 327–335 (1993).
- Mackall, C. & Gress, R. Pathways of T-cell regeneration in mice and humans: implications for bone marrow transplantation and immunotherapy. *Immunol. Rev.* **157**, 61–72 (1997).
- Mackall, C. *et al.* Age, thymopoiesis, and CD4⁺ T-lymphocyte regeneration after intensive chemotherapy. *New Engl. J. Med.* **332**, 143–149 (1995).
- Autran, B. *et al.* Positive effects of combined antiretroviral therapy on CD4⁺ T cell homeostasis and function in advanced HIV disease. *Science* **277**, 112–116 (1997).
- Zhang, Z.-Q. *et al.* Kinetics of CD4⁺ T cell repopulation of lymphoid tissues after treatment of HIV-1 infection. *Proc. Natl Acad. Sci. USA* **95**, 1154–1159 (1998).
- Haynes, B. *et al.* Analysis of the role of the adult thymus in reconstitution of peripheral T lymphocytes in human immunodeficiency virus type 1 infection. *J. Clin. Invest.* (submitted).
- McCune, J. *et al.* High prevalence of thymic tissue in adults with human immunodeficiency virus-1 infection. *J. Clin. Invest.* **101**, 2301–2308 (1998).
- Dybul, M., Kinter, A., Ruiz, M. & Fauci, A. Promethan thymus? *J. Clin. Invest.* **101**, 2299–2300 (1998).
- Kong, F.-K., Chen, C.-L. & Cooper, M. Thymic function can be accurately monitored by the level of recent T cell emigrants in the circulation. *Immunology* **8**, 97–104 (1998).
- Livak, F. & Schatz, D. T-cell receptor α locus V(D)J recombination by-products are abundant in thymocytes and mature T cells. *Mol. Cell. Biol.* **16**, 609–618 (1996).
- Takeshita, S., Toda, M. & Ymagishi, H. Excision products of the T cell receptor gene support a progressive rearrangement model of the α/δ locus. *EMBO J.* **8**, 3261–3270 (1989).
- Bogue, M. & Roth, D. B. Mechanism of V(D)J recombination. *Curr. Opin. Immunol.* **8**, 175–180 (1996).
- Verschuren, M. *et al.* Preferential rearrangements of the T cell receptor- δ -deleting elements in human T cells. *J. Immunol.* **158**, 1208–1216 (1997).
- Pickler, L. *et al.* Control of lymphocyte recirculation in man. *J. Immunol.* **150**, 1105–1121 (1993).
- Petrie, H., Livak, F., Burtrum, D. & Mazel, S. T cell receptor gene recombination patterns and mechanisms: cell death, rescue, and T cell production. *J. Exp. Med.* **182**, 121–127 (1995).
- Piatk, M. Jr *et al.* High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* **259**, 1749–1754 (1993).
- Cossarizza, A. *et al.* Age-related imbalance of virgin (CD45RA⁺) and memory (CD45RO⁺) cells between CD4⁺ and CD8⁺ T lymphocytes in humans: a study from newborns to centenarians. *J. Immunol. Res.* **4**, 118–126 (1992).
- McLean, A. & Michie, C. *In vivo* estimates of division and death rates of human T lymphocytes. *Proc. Natl Acad. Sci. USA* **92**, 3707–3711 (1995).
- Schnittman, S. *et al.* Preferential infection of CD4⁺ memory T cells by human immunodeficiency virus type 1: evidence for a role in the selective T-cell functional defects observed in infected individuals. *Proc. Natl Acad. Sci. USA* **87**, 6058–6062 (1990).
- Chun, T., Chadwick, K., Margolick, J. & Siliciano, R. Differential susceptibility of naive and memory CD4⁺ T cells to the cytopathic effects of infection with human immunodeficiency virus type 1 strain LAI. *J. Virol.* **71**, 4436–4444 (1997).
- Schnittman, S. M. *et al.* Evidence for susceptibility of intrathymic T cell precursors to human immunodeficiency virus infection: a mechanism for T4 (CD4) lymphocyte depletion. *Trans. Assoc. Am. Physicians* **103**, 96–101 (1990).
- Wykrzykowska, J. *et al.* Early regeneration of thymic progenitors in rhesus macaques infected with simian immunodeficiency virus. *J. Exp. Med.* **187**, 1767–1778 (1998).
- Bonyhadi, M. *et al.* HIV induces thymus depletion *in vivo*. *Nature* **363**, 728–732 (1993).
- Aldrovandi, G. M. *et al.* The SCID-hu mouse as a model for HIV-1 infection. *Nature* **363**, 732–736 (1993).
- Pakker, N. *et al.* Biphasic kinetics of peripheral blood T cells after triple combination therapy in HIV-1 infection: a composite of redistribution and proliferation. *Nature Med.* **4**, 208–214 (1998).
- Gorochov, G. *et al.* Perturbation of CD4⁺ and CD8⁺ T-cell repertoires during progression to AIDS and regulation of the CD4⁺ repertoire during antiviral therapy. *Nature Med.* **4**, 215–221 (1998).

29. Withers-Ward, E. *et al.* Transient renewal of thymopoiesis in HIV-infected human thymic implants following antiviral therapy. *Nature Med.* 3, 1102–1109 (1997).
30. Han, S. *et al.* V(D)J recombinase activity in a subset of germinal center B lymphocytes. *Science* 278, 301–305 (1997).

Acknowledgements. We thank J. Wong for monoclonal antibody 12F6; M. Gately for rHL-2; F. Scott, J. McKinsey, A. Rahimi, S. Norris, G. Sempowski, J. Tomasch and M. Zupancic for sample collection and processing; A. Mobley and B. Darnell for FACS support; B. Dawson for viral load measurement; R. Scheuermann and D. Sadora for advice; and all the study participants for their cooperation. This work was supported by grants from the NIH and the American Foundation for AIDS Research. R.A.K., M.B.F. and J.A.Z. are Elizabeth Glaser Scientists of the Pediatric AIDS Foundation.

Correspondence and requests for materials should be addressed to R.A.K. (e-mail: rkoup@mednet.swmed.edu).

An asparaginyl endopeptidase processes a microbial antigen for class II MHC presentation

Bénédicte Manoury*, Eric W. Hewitt*, Nick Morrice†, Pam M. Dando‡, Alan J. Barrett‡ & Colin Watts*

* Department of Biochemistry, Wellcome Sciences Building, and † MRC Protein Phosphorylation Unit, Department of Biochemistry, University of Dundee, Dundee DD1 5EH, UK

‡ Medical Research Council Peptidase Laboratory, Department of Immunology, The Babraham Institute, Babraham Hall, Babraham CB2 4AT, UK

Foreign protein antigens must be broken down within endosomes or lysosomes to generate suitable peptides that will form complexes with class II major histocompatibility complex molecules for presentation to T cells. However, it is not known which proteases are required for antigen processing. To investigate this, we exposed a domain of the microbial tetanus toxin antigen (TTCF) to disrupted lysosomes that had been purified from a human B-cell line. Here we show that the dominant processing activity is not one of the known lysosomal cathepsins, which are generally believed to be the principal enzymes involved in antigen processing, but is instead an asparagine-specific cysteine endopeptidase. This enzyme seems similar or identical to a mammalian homologue¹ of the legumain/haemoglobinase asparaginyl endopeptidases found originally in plants² and parasites³. We designed competitive peptide inhibitors of B-cell asparaginyl endopeptidase (AEP) that specifically block its proteolytic activity and inhibit processing of TTCF *in vitro*. *In vivo*, these inhibitors slow TTCF presentation to T cells, whereas preprocessing of TTCF with AEP accelerates its presentation, indicating that this enzyme performs a key step in TTCF processing. We also show that N-glycosylation of asparagine residues blocks AEP action *in vitro*. This indicates that N-glycosylation could eliminate sites of processing by AEP in mammalian proteins, allowing preferential processing of microbial antigens.

To analyse processing of a foreign antigen without making assumptions about which proteases are involved, we incubated a domain of the tetanus toxin antigen (TTCF; relative molecular mass (M_r) 47 K) with disrupted lysosomes isolated from the human Epstein–Barr virus (EBV)-transformed B-cell line EDR. The antigen was fragmented to produce a discrete series of products at mildly acidic pH, consistent with an endosomal/lysosomal origin for the protease(s) involved (Fig. 1a). Surprisingly, we could not inhibit lysosomal digestion of TTCF with leupeptin or E64 (trans-epoxysuccinyl-L-leucylamido-(4-guanidino)butane), which are both broad-spectrum inhibitors of lysosomal cysteine proteases, or with pepstatin, an inhibitor of the aspartic acid cathepsins E and D (Fig. 1a). In addition, we could not reproduce the lysosomal TTCF-digestion pattern or indeed generate any of these fragments using purified cathepsins L, S, B, D or E (data not shown), all of which have been implicated in the processing of antigens for

presentation on class II major histocompatibility complex (MHC) molecules^{4–7} (reviewed in refs 8–10). However, further inhibitor studies showed that the activity was sensitive to iodoacetamide, N-ethylmaleimide and high concentrations of the diazomethane Z-Phe-Phe-CHN₂ (Fig. 1a and data not shown), indicating that one or more cysteine protease(s) was involved.

We sequenced the major TTCF-digestion products (Fig. 1b) to gain more information on this processing activity. The five major fragments arose from three cleavages in the TTCF protein after residues 873, 1,184 and 1,219 (tetanus toxin numbering), all of which are asparagine residues (Fig. 1b). These results indicated that B-cell lysosomes contain one or more unusual cysteine endopeptidase(s) with possible specificity for asparagine. Although not previously described in antigen-presenting cells, an enzyme called legumain or haemoglobinase, which possesses similar properties, has been found in the seeds of leguminous plants², in *Schistosoma mansoni*³ and in mammalian tissues such as kidney and placenta¹. We partially purified the TTCF-processing activity from a crude B-cell lysosome fraction to establish whether or not it was an asparaginyl endopeptidase. A peak of TTCF-processing activity co-eluted from a cation-exchange resin (~0.4 M NaCl) with an activity capable of cleaving the peptide substrate Z-Val-Ala-Asn-7-(4-methyl)coumaryl-

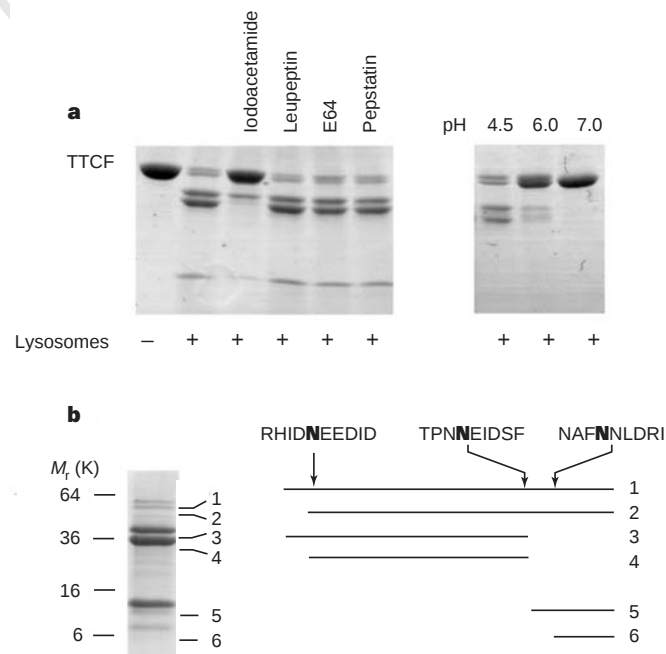


Figure 1 Processing of TTCF by a leupeptin-insensitive cysteine endopeptidase activity. **a**, TTCF was digested *in vitro* in the presence of disrupted lysosomes (1–2 μ g) in the presence or absence of iodoacetamide (1 mM), leupeptin, E64 or pepstatin (each at 0.1 mg ml⁻¹). Digestions were in 0.1 M ammonium acetate at pH 4.5, or Na₂HPO₄ at pH 6.0 or 7.0. After 4 h at 37 °C, the reactions were separated by Tris-tricine SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and stained with Coomassie blue. **b**, Cleavage sites. TTCF (labelled 1) and digestion products (numbered 2–6) were separated by Tris-tricine SDS–PAGE, electrophoretically transferred to PVDF membrane and subjected to five cycles of Edman degradation. The N-terminal sequence obtained for each fragment is shown after the asparagine residue printed in bold. Cleavage occurs after asparagines 873 (fragments 2 and 4), 1,184 (fragment 5) and 1,219 (fragment 6) numbered from the whole toxin sequence. Note that the first authentic tetanus toxin residue is Asp 872. There are 24 N-terminal residues that derive from the histidine tagging of TTCF (see Methods). The N-terminal sequence of His-tagged TTCF was obtained for fragments 1 and 3.

amide (Fig. 2a, b). Other protease activities, capable of cleaving the substrates Z-Phe-Arg-NHMec (cathepsin L/B substrate) and Z-Val-Val-Arg-NHMec (cathepsin S substrate), partly overlapped the asparaginyl endopeptidase activity (Fig. 2a). However, these activities were completely inhibited by E64, whereas both the TTCF- and the Z-Val-Ala-Asn-NHMec-processing activities were unaffected (data not shown), confirming that a new cysteine endopeptidase was involved. To compare this activity to the mammalian form of legumain, we generated antisera against peptides based on the human legumain complementary DNA sequence¹. An antiserum against the peptide KGIGSGKVLKSGPQ crossreacted with purified legumain from pig kidney and with a human B-cell lysosome protein of similar M_r (~35K) that co-eluted with the TTCF- and peptide-substrate-processing activity (Fig. 2c). Reactivity with this protein was blocked specifically by the peptide used as an immunogen (Fig. 2c). Finally, the same processing pattern and cleavage sites were found when TTCF was cleaved either by the partially purified human B-cell asparaginyl endopeptidase or by the purified legumain from pig kidney (Figs 1b and 2d, ref. 1 and data not shown). Thus, EBV-transformed B lymphocytes contain an asparaginyl endopeptidase similar or identical to human legumain and this dominates processing of the native form of the TTCF antigen *in vitro*. We use the acronym AEP to describe the human B-cell enzyme.

As there are currently no specific inhibitors available for asparaginyl endopeptidases, we tested the possibility that asparagine-containing peptides might act as competitive inhibitors. The amino- and carboxy-terminally blocked tetrapeptide Fmoc-Ala-Glu-Asn-Lys-NH₂ (AENK) caused substantial inhibition of TTCF

processing by pig kidney legumain (Fig. 3a, left panel). Another blocked peptide, Fmoc-Lys-Asn-Asn-Glu-NH₂ (KNNE), also inhibited processing but was less effective. Processing of TTCF by B-cell AEP was blocked almost completely by 1 mg ml⁻¹ AENK (Fig. 3a, right panel). Importantly, the same concentrations of the glutamine analogues Fmoc-Ala-Glu-Gln-Lys-NH₂ (AEQK) and Fmoc-Lys-Gln-Gln-Glu-NH₂ (KQQE) had no effect on TTCF processing by asparaginyl endopeptidases (Fig. 3a). To establish that AENK blocked AEP but not other proteases, we measured the hydrolysis, by B-cell lysosomal extracts, of different synthetic peptide substrates in the presence and absence of these tetrapeptides. Increasing concentrations of AENK blocked cleavage of Z-Val-Ala-Asn-NHMec by B-cell lysosomes (Fig. 3b, c), but had no effect on hydrolysis of the cathepsin S substrate Z-Val-Val-Arg-NHMec (Fig. 3d) or the cathepsin B/L substrate Z-Phe-Arg-NHMec (Fig. 3e). In contrast, leupeptin inhibited hydrolysis of the cathepsin S and B/L substrates (Fig. 3d and data not shown), but not Z-Val-Ala-Asn-NHMec (Fig. 3c).

If AEP is important in TTCF processing *in vivo*, then the competing AENK peptide might be expected to interfere with normal loading of TTCF epitopes onto class II MHC molecules. To test this, freshly isolated peripheral blood mononuclear cells (PBMCs) were pre-incubated briefly with AENK, or the control peptide AEQK, and then pulsed with TTCF antigen in the continued presence of one or the other peptide. At different times, the cells were washed, fixed and then co-cultured with different autologous T-cell clones to assess expression of different peptide-MHC complexes. Presentation was detectable within 30 or 60 min of antigen pulsing in the presence of the control peptide (Fig. 4a),

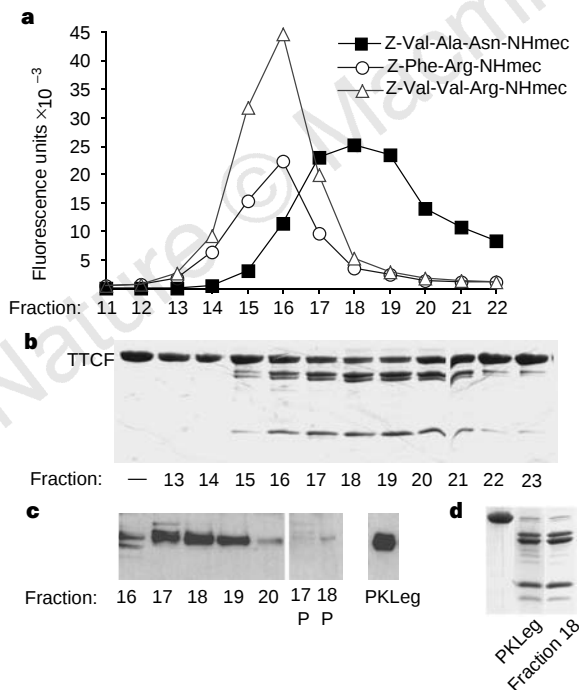


Figure 2 TTCF is processed by an asparaginyl endopeptidase. **a**, **b**, A B-cell lysosomal extract was applied to a cation-exchange resin (Mono-S) and eluted with a 0.01–1.0 M NaCl gradient. Each fraction was incubated **a**, with the substrates Z-Val-Ala-Asn-NHMec, Z-Phe-Arg-NHMec or Z-Val-Val-Arg-NHMec or **b**, with 10 μ g TTCF for 4 h at 37 °C before Tris-tricine SDS-PAGE analysis. **c**, Each fraction was analysed by Tris-tricine SDS-PAGE and western blotting with an affinity-purified sheep antiserum in the presence (17P and 18P) or absence of the immunizing peptide (KGIGSGKVLKSGPQ) from human legumain¹. **d**, Comparison of TTCF processing with either partially purified B-cell asparaginyl endopeptidase (fraction 18 from **b** above) or purified legumain from pig kidney.

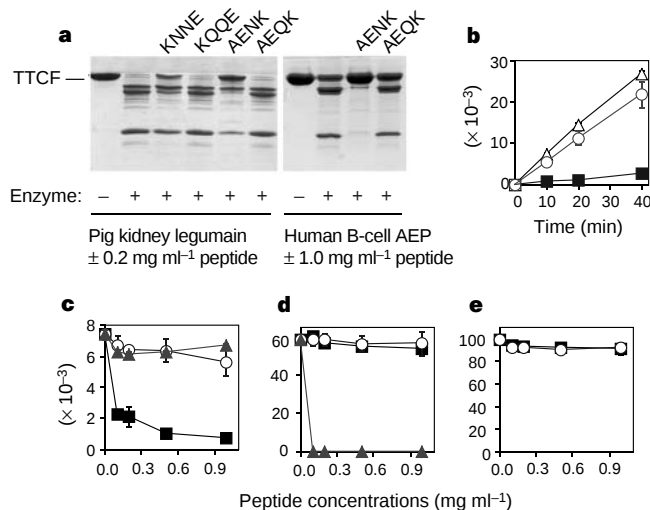


Figure 3 Specific peptide inhibitors of asparaginyl endopeptidase. **a**, Specific inhibition of TTCF processing by asparagine peptides. TTCF (10 μ g) was digested in 20 μ l either with purified legumain from pig kidney (0.6 μ g ml⁻¹) or with human B-cell AEP in the presence or absence of the N- and C-terminally blocked peptides indicated. **b–e**, Specific inhibition of AEP activity by Fmoc-Ala-Glu-Asn-Lys-NH₂ (AENK). **b**, A crude lysosomal protease mixture was incubated with Z-Val-Ala-Asn-NHMec in the absence (open triangles) or presence of AENK (filled squares) or AEQK (open circles) and release of 7-amino-4-methyl coumarin (NHMec) was measured at 460 nm after 10, 20 or 40 min. **c–e**, Hydrolysis of Z-Val-Ala-Asn-NHMec (**c**), Z-Val-Val-Arg-NHMec (**d**) or Z-Phe-Arg-NHMec (**e**) was measured after a 10-min incubation in the presence of increasing concentrations of AENK (filled squares), AEQK (open circles) or leupeptin (filled triangles). Leupeptin was tested in **c**, **d** only.

but in the presence of AENK presentation to most clones (for example, AK 5, 90 and 6) was undetectable at these early times. We also generated an EBV-transformed B-cell line from donor A.K. and tested the effect of the tetrapeptides on B-cell antigen presentation to the same T-cell clones. The AENK peptide blocked antigen presentation by EBV-B cells very efficiently, whereas the control peptide AEQK had virtually no effect on the kinetics of antigen presentation (Fig. 4b). Thus AENK specifically inhibited antigen presentation by two different populations of antigen-presenting cell.

Antigen presentation to all clones recovered to different extents after longer times of antigen pulsing. We suggest that this recovery is expected given that AENK inhibits by acting as a competitive substrate and is itself hydrolysed by asparaginyl endopeptidases (data not shown). Under these conditions it is not possible to block TCTF processing completely. We also observed that some clones, for example clone 33, were less sensitive to blockade of AEP (Fig. 4a, b). Clonal variations such as these might be explained by a differential requirement for AEP processing, dependent on the location of the T-cell epitope, the antigen dose-response of individual clones, or a combination of these factors. Consistent with the first scenario, we have observed that clones specific for the same peptide region do show similar sensitivities to AEP blockade. For example, clones 5 and 90 and a third clone (AK 20) all recognize the peptide region 1,145–1,156 and show similar sensitivity to AENK (Fig. 4 and data not shown). However, as expected, there was also variation in the antigen dose-response among the clones we tested. Those requiring higher levels of antigen for half-maximal triggering showed, in general, slower kinetics of presentation and greater sensitivity to AEP blockade. We suspect that the apparent differential requirement for AEP processing shown by different clones is due to a combination of epitope location, clone 'affinity' and possibly other factors. Nonetheless, it is clear that presentation of most tetanus toxin T-cell epitopes that have been tested so far is affected by AEP blockade. This is consistent with the hypothesis that initial processing by AEP may be the 'key' that 'unlocks' the native protein for further processing.

As a further test of this hypothesis, we assayed the kinetics of

presentation of TCTF that had been predigested *in vitro* with AEP, reasoning that this might accelerate antigen presentation. Cells incubated with TCTF preprocessed with AEP presented T-cell epitopes after only 15–30 min, whereas unprocessed TCTF required 60–120 min before presentation could be detected (Fig. 4c). AENK did not block the rapid presentation of the preprocessed antigen at these times (Fig. 4d), confirming that this inhibitor blocks AEP action specifically on the antigen substrate *in vivo*. However, TCTF preprocessed with AEP still required further cellular processing as it could not be presented by fixed PBMCs (Fig. 4c, d). Taken together, these results support a working model in which AEP initiates TCTF proteolysis but generation of suitable peptides for MHC binding requires further intracellular processing. Native TCTF was a poor substrate for cathepsins E, D, L, S and B (data not shown), but many processing sites, for example for cathepsins E and D, were revealed after antigen unfolding⁷.

We did not expect a highly specific enzyme, not previously known to exist in antigen-presenting cells, to dominate processing of the native TCTF antigen *in vitro* and to control the kinetics of its presentation *in vivo*. Other protein 'antigens', such as ovalbumin, RNase and hen egg lysozyme, are cleaved by AEP and we have detected this activity in different human and murine antigen-presenting cells (data not shown). To our knowledge, this is the first time that a specific protease has been shown to have a central role in processing a defined antigen. The specificity of this enzyme for asparagine indicated that it might have additional significance. Specifically, we wondered whether N-glycosylation of asparagine residues might interfere with AEP action on protein and peptide substrates. We tested this hypothesis using both synthetic and natural glycopeptide substrates. The sequence HIDNEEDI, which contains one of the three asparaginyl bonds cleaved by both pig kidney legumain and B-cell AEP (Fig. 1), was synthesized with and without modification of the asparagine residue with N-glucosamine. Because the asparagine in this sequence is not part of a consensus glycosylation site (N-X-S/T), we also tested the peptide HIDNESDI, in which Glu in the P2' position is changed to Ser. Each peptide was incubated with legumain from pig kidney and analysed using high-performance liquid chromatography (HPLC). Whereas

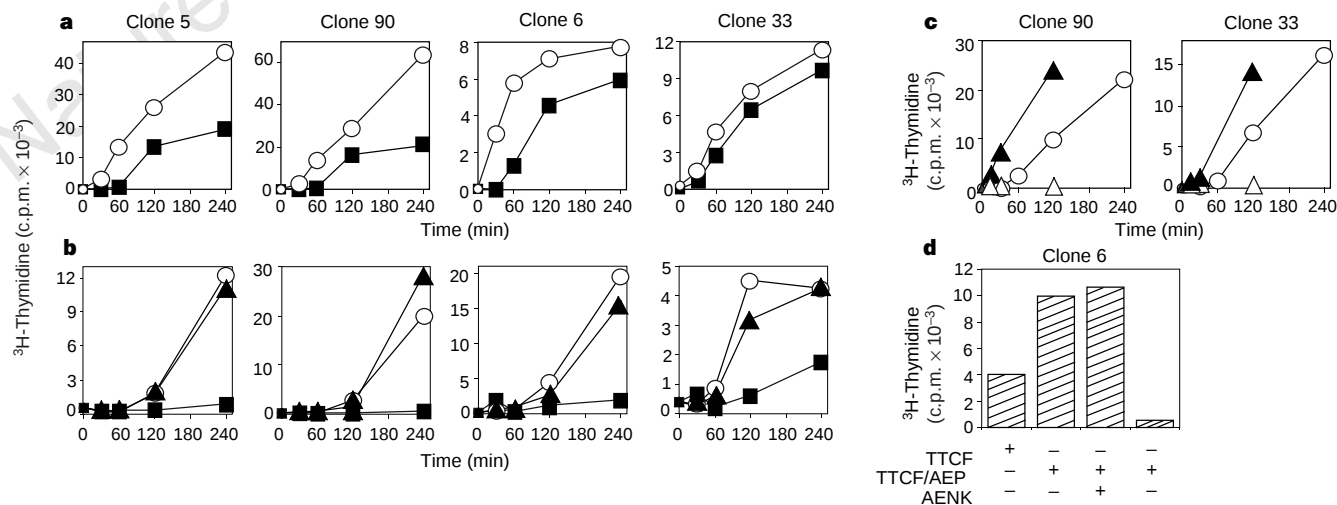


Figure 4 TCTF processing and presentation requires AEP *in vivo*. **a, b**, Peptide inhibitors of asparaginyl endopeptidase slow antigen presentation *in vivo*. **a**, Peripheral blood mononuclear cells or **b**, EBV-transformed B-cells, both from donor A.K., were pre-incubated for 15 min in the presence of AENK (filled squares), AEQK (open circles) or no peptide (filled triangles), before addition of TCTF antigen. When present, peptide was maintained during antigen pulsing at 1 mg ml⁻¹ (**a**) or 2 mg ml⁻¹ (**b**). At different times the cells were washed and fixed in 0.05% glutaraldehyde and antigen presentation was measured using proliferation

of different autologous TCTF-specific T-cell clones. **c**, AEP preprocessing of TCTF speeds antigen presentation. Antigen presentation was measured as in **a**, but using either TCTF (open circles) or TCTF predigested with AEP (filled and open triangles) and either live (open circles and filled triangles) or fixed (open triangles) autologous PBMCs. **d**, Presentation of AEP-preprocessed antigen is no longer inhibited by AENK. Presentation was measured after 60 min of antigen pulsing using EBV-transformed B cells.

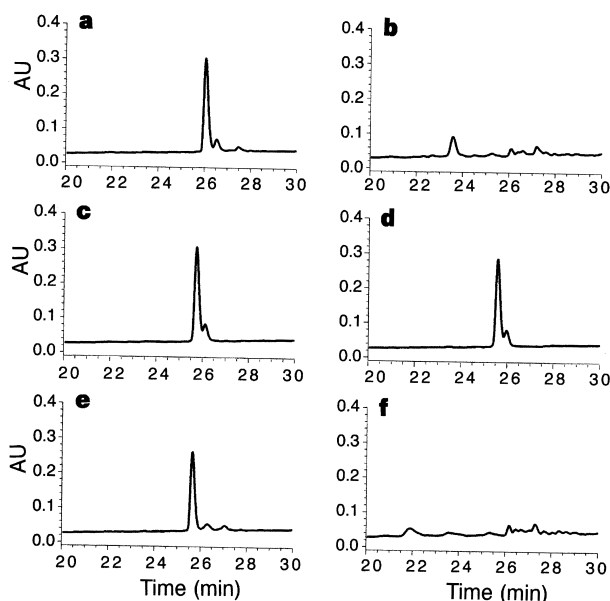


Figure 5 Asparagine *N*-glycosylation blocks asparaginyl endopeptidase action. The TTCF-based peptides HIDNEEDI (**a**, **b**), HIDN(*N*-glucosamine)EEDI (**c**, **d**) and HIDNESDI (**e**, **f**) were incubated in the presence (**b**, **d**, **f**) or absence (**a**, **c**, **e**) of 50 mU ml⁻¹ pig kidney legumain in sodium citrate/phosphate buffer at 30 °C before analysis by reverse-phase HPLC.

the control peptide and the one containing the consensus glycosylation site were completely hydrolysed (Fig. 5a, b, e, f), the HIDN(*N*-Glc)EEDI peptide was completely resistant (Fig. 5c, d). We also analysed a natural 21-residue tryptic glycopeptide (residues 622–642) purified from human serum transferrin (see Table 1). This peptide contains two asparagine residues, one of which (N630) carries a biantennary oligosaccharide¹¹. The same peptide was synthesized without modification of the asparagine and digested in parallel. (Three lysine residues were added to improve solubility.) Both mass spectrometric analysis and N-terminal sequencing of the products confirmed that hydrolysis of both asparagines 630 and 637 occurred in the non-glycosylated peptide (Table 1). However, we

detected products from hydrolysis of only N637 in the glycopeptide (Table 1). A second transferrin-derived tryptic glycopeptide (residues 421–452), which contained three asparagine residues, was also digested with AEP. Again, we detected fragments from hydrolysis only at the non-glycosylated asparagine residues 430 and 436 (not shown).

Taken together, these data show that *N*-glycosylation blocks attack by asparaginyl endopeptidase at otherwise sensitive sites. Thus it is possible that non-glycosylated microbial proteins such as TTCF will, in general, be more sensitive to processing by AEP. The innate immune system's ability to recognize non-self carbohydrate structures^{12,13} might also extend, in the case of prokaryotic microbial proteins, to an 'innate' bias towards processing of non-*N*-glycosylated microbial antigens. Although we do not suggest that glycoproteins are completely resistant to AEP processing, strategic *N*-glycosylation may alter the pathway of proteolysis and unfolding so as to disfavor loading of particular epitopes. Viruses and other pathogens might exploit this to avoid presentation of otherwise immunogenic peptides. Thus, *N*-glycosylation may provide viral glycoproteins with both a physical barrier, for example to neutralizing antibodies¹⁴, and a means of limiting processing by AEP in the class II MHC pathway. This may be one reason why T-cell immune responses can be raised to non-glycosylated recombinant viral glycoproteins that are not recalled by native, fully glycosylated viral glycoprotein¹⁵. Finally, it is conceivable that disturbances of *N*-glycosylation might result in altered processing of self proteins by AEP, leading to presentation of 'cryptic' epitopes with possible pathological consequences. □

Methods

TTCF-processing activity. Lysosomal fractions from the human B-cell line EDR were prepared on 27% Percoll density gradients as described previously¹⁶ using ¹²⁵I-labelled transferrin (internalized for 30 min at 37 °C) and β -hexosaminidase to identify endosomal/plasma membrane and lysosomal fractions, respectively. Lysosomal membranes were collected by centrifugation at 23,000g for 1 h, removed from the underlying Percoll pellet, concentrated at 50,000g, and stored frozen. For separation of leupeptin-insensitive cysteine protease activity, 7.5×10^9 cells were homogenized in a ball-bearing homogenizer and nuclei and unbroken cells were removed by centrifugation at 2,000g for 10 min. A membrane pellet was collected at 40,000g and solubilized in 20 mM sodium acetate, 1 mM EDTA, pH 5.2, containing 0.1% CHAPS. The extract was centrifuged at 2,000g for 20 min and then applied to a Mono-S column (Pharmacia) equilibrated in 20 mM sodium acetate, pH 5.2, 10 mM NaCl. Fractions (1 ml) were eluted in 20 mM sodium acetate, pH 5.2, first with a gradient of NaCl (10–300 mM in 5 ml), followed by a 3-ml wash at 300 mM NaCl and then a second NaCl gradient (300–500 mM in 8 ml; 500–1,000 mM in 4 ml). Each fraction was tested for TTCF-processing activity and ability to cleave various fluorogenic peptide substrates. Fluorometric protease assays were done as described^{1,17}.

Proteins and peptides. A histidine-tagged derivative of the C-terminal domain of tetanus toxin was prepared and purified as described^{7,18}. This protein includes residues 872–1,315 of the complete toxin (1–1,315 residues) preceded by the sequence MGHGHHHHHHHHHHSSGHIEGRHI. Tetrapeptides were synthesized using Fmoc chemistry leaving the C terminus amidated and the N terminus retaining the Fmoc group. The peptide KGIGSGKVLKSGPQ from the human preprolegumain sequence¹ was coupled by an additional C-terminal cysteine to KLH and BSA and used to immunize a sheep (Scottish Antibody Production Unit, Carlisle). Peptide-specific antibodies obtained after three boosts were purified as described¹⁹. Z-Val-Ala-Asn-NHMe was prepared as described²; other substrates were obtained from Sigma or Bachem.

Antigen presentation. PBMCs were prepared by Ficoll/Paque centrifugation and used fresh. T-cell clones specific for TTCF were established from donor A.K. as described²⁰. Epitope mapping was done using a set of 88 peptides, 17 residues in length and spanning the TTCF sequence (Chiron Mimotopes). Clones AK 6, 90 and 33 recognize regions 1,235–1,246, 1,145–1,156 and 1,125–1,136, respectively. Clone 5 also recognizes the region 1,145–1,156. An EBV-transformed cell line from the same donor was established by standard

Table 1 *N*-glycosylation blocks AEP action on a natural glycopeptide

	Predicted mass	Observed mass
Synthetic transferrin peptide (622–642KKK)		
No enzyme		
QQQHLFGSNVTDCSGNFCLFRKKK	2,784.36	2,784.17
+ Asparaginyl endopeptidase		
QQQHLFGSN	1,058.51	1,058.46
VTDCSGNFCLFRKKK	1,745.83	1,743.69*
FCLFRKKK	1,069.63	1,069.59
Transferrin glycopeptide (622–642)		
No enzyme		
QQQHLFGSN(CHO)VTDCSGNFCLFR	4,720.76	4,720.2
+ Asparaginyl endopeptidase		
QQQHLFGSN(CHO)VTDCSGN	3,996.88	3,996.85
FCLFR	742.91	742.23
VTDCSGNFCLFR	1,476.66	Not observed
QQQHLFGSN(CHO)	3,263.13	Not observed

Predicted fragment masses were obtained from the human transferrin sequence (Swiss-Prot accession number: P02787 calculated using PeptideMass at the EXPASY server). The transferrin biantennary oligosaccharide (2 sialic acid + 4 *N*-acetylglucosamine + 5 hexose) has a mass of 2,204 coupled to peptide¹¹. Variants also observed but not shown are monosialylated forms of the glycopeptides and N-terminal peptide forms that are 17 mass units lower corresponding to N-terminal glutamine pyroglutamination.

* Cyclization of the two cysteine residues in this peptide probably accounts for the observed mass deficit of 2.

methods. Peptides AENK or AEQK were dissolved in water, made isotonic with NaCl and diluted into RPMI growth medium. T-cell-proliferation assays were done essentially as described^{20,21}. Briefly, after antigen pulsing ($30 \mu\text{g ml}^{-1}$ TTCF) with tetrapeptides ($1\text{--}2 \text{ mg ml}^{-1}$), PBMCs or EBV-B cells were washed in PBS and fixed for 45 s in 0.05% glutaraldehyde. Glycine was added to a final concentration of 0.1M and the cells were washed five times in RPMI 1640 medium containing 1% FCS before co-culture with T-cell clones in round-bottom 96-well microtitre plates. After 48 h, the cultures were pulsed with $1 \mu\text{Ci}$ of ^3H -thymidine and harvested for scintillation counting 16 h later. Predigestion of native TTCF was done by incubating $200 \mu\text{g}$ TTCF with $0.25 \mu\text{g}$ pig kidney legumain in $500 \mu\text{l}$ 50 mM citrate buffer, pH 5.5, for 1 h at 37°C .

Glycopeptide digestions. The peptides HIDNEEDI, HIDN(N-glucosamine) EEDI and HIDNESDI, which are based on the TTCF sequence, and QQQHLFGSNVTDCSGNFCLFR(KKK), which is based on human transferrin, were obtained by custom synthesis. The three C-terminal lysine residues were added to the natural sequence to aid solubility. The transferrin glycopeptide QQQHLFGSNVTDCSGNFCLFR was prepared by tryptic (Promega) digestion of 5 mg reduced, carboxy-methylated human transferrin followed by concanavalin A chromatography¹¹. Glycopeptides corresponding to residues 622–642 and 421–452 were isolated by reverse-phase HPLC and identified by mass spectrometry and N-terminal sequencing. The lyophilized transferrin-derived peptides were redissolved in 50 mM sodium acetate, pH 5.5, 10 mM dithiothreitol, 20% methanol. Digestions were performed for 3 h at 30°C with $5\text{--}50 \text{ mU ml}^{-1}$ pig kidney legumain or B-cell AEP. Products were analysed by HPLC or MALDI-TOF mass spectrometry using a matrix of 10 mg ml^{-1} α -cyanocinnamic acid in 50% acetonitrile/0.1% TFA and a PerSeptive Biosystems Elite STR mass spectrometer set to linear or reflector mode. Internal standardization was obtained with a matrix ion of 568.13 mass units.

Received 29 September; accepted 3 November 1998.

- Chen, J. M. *et al.* Cloning, isolation, and characterisation of mammalian leucine aminopeptidase. *J. Biol. Chem.* **272**, 8090–8098 (1997).
- Kembhavi, A. A., Buttle, D. J., Knight, C. G. & Barrett, A. J. The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluoromethylating assays. *Arch. Biochem. Biophys.* **303**, 208–213 (1993).
- Dalton, J. P., Hala Jmriska, L. & Bridley, P. J. Asparaginyl endopeptidase activity in adult *Schistosoma mansoni*. *Parasitology* **111**, 575–580 (1995).
- Bennett, K. *et al.* Antigen processing for presentation by class II major histocompatibility complex requires cleavage by cathepsin E. *Eur. J. Immunol.* **22**, 1519–1524 (1992).
- Riese, R. J. *et al.* Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading. *Immunity* **4**, 357–366 (1996).
- Rodriguez, G. M. & Diment, S. Role of cathepsin D in antigen presentation of ovalbumin. *J. Immunol.* **149**, 2894–2898 (1992).
- Hewitt, E. W. *et al.* Natural processing sites for human cathepsin E and cathepsin D in tetanus toxin: implications for T cell epitope generation. *J. Immunol.* **159**, 4693–4699 (1997).
- Watts, C. Capture and processing of exogenous antigens for presentation on MHC molecules. *Annu. Rev. Immunol.* **15**, 821–850 (1997).
- Chapman, H. A. Endosomal proteases and MHC class II function. *Curr. Opin. Immunol.* **10**, 93–102 (1998).
- Fineschi, B. & Miller, J. Endosomal proteases and antigen processing. *Trends Biochem. Sci.* **22**, 377–382 (1997).
- Lu, J. & van Halbeek, H. Complete ^1H and ^{13}C resonance assignments of a 21-amino acid glycopeptide prepared from human serum transferrin. *Carbohydr. Res.* **296**, 1–21 (1996).
- Fearon, D. T. & Locksley, R. M. The instructive role of innate immunity in the acquired immune response. *Science* **272**, 50–54 (1996).
- Medzhitov, R. & Janeway, C. A. J. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* **91**, 295–298 (1997).
- Wyatt, R. *et al.* The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393**, 705–711 (1998).
- Botarelli, P. *et al.* N-glycosylation of HIV gp120 may constrain recognition by T lymphocytes. *J. Immunol.* **147**, 3128–3132 (1991).
- Davidson, H. W., West, M. A. & Watts, C. Endocytosis, intracellular trafficking, and processing of membrane IgG and monovalent antigen/membrane IgG complexes in B lymphocytes. *J. Immunol.* **144**, 4101–4109 (1990).
- Barrett, A. J. & Kirschke, H. Cathepsin B, cathepsin H and cathepsin L. *Methods Enzymol.* **80**, 535–559 (1981).
- Makoff, A. J., Ballantine, S. P., Smallwood, A. E. & Fairweather, N. F. Expression of tetanus toxin fragment C in *E. coli*: its purification and potential use as a vaccine. *Biotechnology* **7**, 1043–1046 (1989).
- Lane, D. P. & Harlow, E. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1988).
- Lanzavecchia, A. Antigen-specific interaction between T and B cells. *Nature* **314**, 537–539 (1985).
- Pond, L. & Watts, C. Characterization of transport of newly assembled, T cell-stimulatory MHC class II-peptide complexes from MHC class II compartments to the cell surface. *J. Immunol.* **159**, 543–553 (1997).

Acknowledgements. We thank M. Ferguson for helpful discussions and advice; E. Smythe and L. Grayson for advice and technical assistance; B. Spruce, A. Knight and the BTS (Ninewells Hospital) for help with blood monocyte preparation; and our colleagues for many helpful comments on the manuscript. This work was supported by the Wellcome Trust and by an EMBO Long-term fellowship to B. M.

Correspondence and requests for materials should be addressed to C.W. (e-mail: c.watts@dundee.ac.uk).

Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer

Robert M. Pitti[†], Scot A. Marsters[†], David A. Lawrence[†], Margaret Roy^{*}, Frank C. Kischkel^{*}, Patrick Dowd^{*}, Arthur Huang^{*}, Christopher J. Donahue^{*}, Steven W. Sherwood^{*}, Daryl T. Baldwin^{*}, Paul J. Godowski^{*}, William I. Wood^{*}, Austin L. Gurney^{*}, Kenneth J. Hillan^{*}, Robert L. Cohen^{*}, Audrey D. Goddard^{*}, David Botstein[‡] & Avi Ashkenazi^{*}

^{*} Departments of Molecular Oncology, Molecular Biology, and Immunology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA

[‡] Department of Genetics, Stanford University, Stanford, California 94305, USA

[†] These authors contributed equally to this work

Fas ligand (FasL) is produced by activated T cells and natural killer cells and it induces apoptosis (programmed cell death) in target cells through the death receptor Fas/Apo1/CD95 (ref. 1). One important role of FasL and Fas is to mediate immune-cytotoxic killing of cells that are potentially harmful to the organism, such as virus-infected or tumour cells¹. Here we report the discovery of a soluble decoy receptor, termed decoy receptor 3 (DcR3), that binds to FasL and inhibits FasL-induced apoptosis. The DcR3 gene was amplified in about half of 35 primary lung and colon tumours studied, and DcR3 messenger RNA was expressed in malignant tissue. Thus, certain tumours may escape FasL-dependent immune-cytotoxic attack by expressing a decoy receptor that blocks FasL.

By searching expressed sequence tag (EST) databases, we identified a set of related ESTs that showed homology to the tumour necrosis factor (TNF) receptor (TNFR) gene superfamily². Using the overlapping sequence, we isolated a previously unknown full-length complementary DNA from human fetal lung. We named the protein encoded by this cDNA decoy receptor 3 (DcR3). The cDNA encodes a 300-amino-acid polypeptide that resembles members of the TNFR family (Fig. 1a): the amino terminus contains a leader sequence, which is followed by four tandem cysteine-rich domains (CRDs). Like one other TNFR homologue, osteoprotegerin (OPG)³, DcR3 lacks an apparent transmembrane sequence, which indicates that it may be a secreted, rather than a membrane-associated, molecule. We expressed a recombinant, histidine-tagged form of DcR3 in mammalian cells; DcR3 was secreted into the cell culture medium, and migrated on polyacrylamide gels as a protein of relative molecular mass 35,000 (data not shown). DcR3 shares sequence identity in particular with OPG (31%) and TNFR2 (29%), and has relatively less homology with Fas (17%). All of the cysteines in the four CRDs of DcR3 and OPG are conserved; however, the carboxy-terminal portion of DcR3 is 101 residues shorter.

We analysed expression of DcR3 mRNA in human tissues by northern blotting (Fig. 1b). We detected a predominant 1.2-kilobase transcript in fetal lung, brain, and liver, and in adult spleen, colon and lung. In addition, we observed relatively high DcR3 mRNA expression in the human colon carcinoma cell line SW480.

To investigate potential ligand interactions of DcR3, we generated a recombinant, Fc-tagged DcR3 protein. We tested binding of DcR3-Fc to human 293 cells transfected with individual TNF-family ligands, which are expressed as type 2 transmembrane proteins (these transmembrane proteins have their N termini in the cytosol). DcR3-Fc showed a significant increase in binding to cells transfected with FasL⁴ (Fig. 2a), but not to cells transfected with TNF⁵, Apo2L/TRAIL^{6,7}, Apo3L/TWEAK^{8,9}, or OPGL/TRANCE/

RANKL¹⁰⁻¹² (data not shown). DcR3-Fc immunoprecipitated shed FasL from FasL-transfected 293 cells (Fig. 2b) and purified soluble FasL (Fig. 2c), as did the Fc-tagged ectodomain of Fas but not TNFR1. Gel-filtration chromatography showed that DcR3-Fc and soluble FasL formed a stable complex (Fig. 2d). Equilibrium analysis indicated that DcR3-Fc and Fas-Fc bound to soluble FasL with a comparable affinity ($K_d = 0.8 \pm 0.2$ and 1.1 ± 0.1 nM, respectively; Fig. 2e), and that DcR3-Fc could block nearly all of the binding of soluble FasL to Fas-Fc (Fig. 2e, inset). Thus, DcR3 competes with Fas for binding to FasL.

To determine whether binding of DcR3 inhibits FasL activity, we tested the effect of DcR3-Fc on apoptosis induction by soluble FasL in Jurkat T leukaemia cells, which express Fas (Fig. 3a). DcR3-Fc and Fas-Fc blocked soluble-FasL-induced apoptosis in a similar dose-dependent manner, with half-maximal inhibition at $\sim 0.1 \mu\text{g ml}^{-1}$. Time-course analysis showed that the inhibition did not merely delay cell death, but rather persisted for at least 24 hours (Fig. 3b). We also tested the effect of DcR3-Fc on activation-induced cell death (AICD) of mature T lymphocytes, a FasL-dependent process¹. Consistent with previous results¹³, activation of interleukin-2-stimulated CD4-positive T cells with anti-CD3 antibody increased the level of apoptosis twofold, and Fas-Fc blocked this effect substantially (Fig. 3c); DcR3-Fc blocked the

induction of apoptosis to a similar extent. Thus, DcR3 binding blocks apoptosis induction by FasL.

FasL-induced apoptosis is important in elimination of virus-infected cells and cancer cells by natural killer cells and cytotoxic T lymphocytes; an alternative mechanism involves perforin and granzymes^{1,14-16}. Peripheral blood natural killer cells triggered marked cell death in Jurkat T leukaemia cells (Fig. 3d); DcR3-Fc and Fas-Fc each reduced killing of target cells from $\sim 65\%$ to $\sim 30\%$, with half-maximal inhibition at $\sim 1 \mu\text{g ml}^{-1}$; the residual killing was probably mediated by the perforin/granzyme pathway. Thus, DcR3 binding blocks FasL-dependent natural killer cell activity. Higher DcR3-Fc and Fas-Fc concentrations were required to block natural killer cell activity compared with those required to block soluble FasL activity, which is consistent with the greater potency of membrane-associated FasL compared with soluble FasL¹⁷.

Given the role of immune-cytotoxic cells in elimination of tumour cells and the fact that DcR3 can act as an inhibitor of FasL, we proposed that DcR3 expression might contribute to the ability of some tumours to escape immune-cytotoxic attack. As genomic amplification frequently contributes to tumorigenesis, we investigated whether the DcR3 gene is amplified in cancer. We analysed DcR3 gene-copy number by quantitative polymerase chain

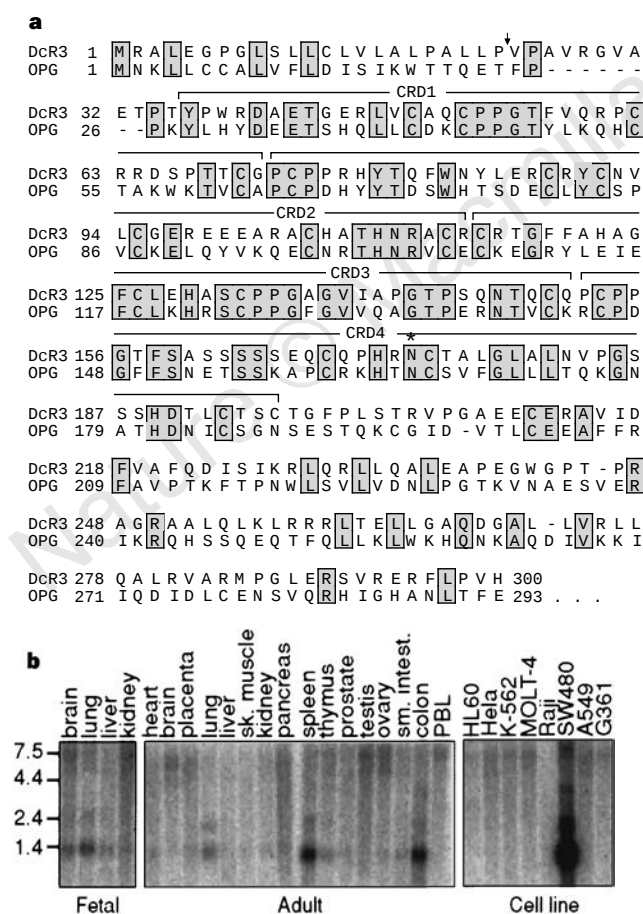


Figure 1 Primary structure and expression of human DcR3. **a**, Alignment of the amino-acid sequences of DcR3 and of osteoprotegerin (OPG); the C-terminal 101 residues of OPG are not shown. The putative signal cleavage site (arrow), the cysteine-rich domains (CRD 1-4), and the N-linked glycosylation site (asterisk) are shown. **b**, Expression of DcR3 mRNA. Northern hybridization analysis was done using the DcR3 cDNA as a probe and blots of poly(A)⁺ RNA (Clontech) from human fetal and adult tissues or cancer cell lines. PBL, peripheral blood lymphocyte.

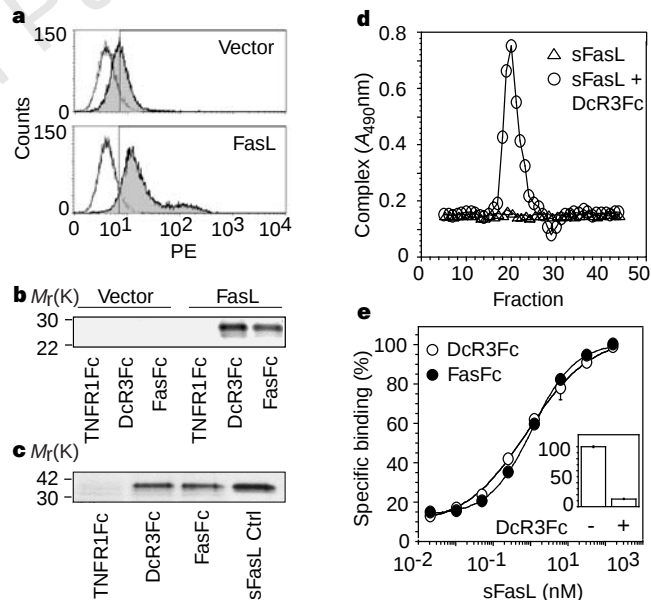


Figure 2 Interaction of DcR3 with FasL. **a**, 293 cells were transfected with pRK5 vector (top) or with pRK5 encoding full-length FasL (bottom), incubated with DcR3-Fc (solid line, shaded area), TNFR1-Fc (dotted line) or buffer control (dashed line) (the dashed and dotted lines overlap), and analysed for binding by FACS. Statistical analysis showed a significant difference ($P < 0.001$) between the binding of DcR3-Fc to cells transfected with FasL or pRK5. PE, phycoerythrin-labelled cells. **b**, 293 cells were transfected as in **a** and metabolically labelled, and cell supernatants were immunoprecipitated with Fc-tagged TNFR1, DcR3 or Fas. **c**, Purified soluble FasL (sFasL) was immunoprecipitated with TNFR1-Fc, DcR3-Fc or Fas-Fc and visualized by immunoblot with anti-FasL antibody. sFasL was loaded directly for comparison in the right-hand lane. **d**, Flag-tagged sFasL was incubated with DcR3-Fc or with buffer and resolved by gel filtration; column fractions were analysed in an assay that detects complexes containing DcR3-Fc and sFasL-Flag. **e**, Equilibrium binding of DcR3-Fc or Fas-Fc to sFasL-Flag. Inset, competition of DcR3-Fc with Fas-Fc for binding to sFasL-Flag.

reaction (PCR)¹⁸ in genomic DNA from 35 primary lung and colon tumours, relative to pooled genomic DNA from peripheral blood leukocytes (PBLs) of 10 healthy donors. Eight of 18 lung tumours and 9 of 17 colon tumours showed DcR3 gene amplification, ranging from 2- to 18-fold (Fig. 4a, b). To confirm this result, we analysed the colon tumour DNAs with three more, independent sets of DcR3-based PCR primers and probes; we observed nearly the same amplification (data not shown).

We then analysed DcR3 mRNA expression in primary tumour tissue sections by *in situ* hybridization. We detected DcR3 expression in 6 out of 15 lung tumours, 2 out of 2 colon tumours, 2 out of 5 breast tumours, and 1 out of 1 gastric tumour (data not shown). A section through a squamous-cell carcinoma of the lung is shown in Fig. 4c. DcR3 mRNA was localized to infiltrating malignant epithelium, but was essentially absent from adjacent stroma, indicating tumour-specific expression. Although the individual tumour specimens that we analysed for mRNA expression and gene amplification were different, the *in situ* hybridization results are consistent with the finding that the DcR3 gene is amplified frequently in tumours. SW480 colon carcinoma cells, which showed abundant DcR3 mRNA expression (Fig. 1b), also had marked DcR3 gene amplification, as shown by quantitative PCR (fourfold) and by Southern blot hybridization (fivefold) (data not shown).

If DcR3 amplification in cancer is functionally relevant, then DcR3 should be amplified more than neighbouring genomic regions that are not important for tumour survival. To test this,

we mapped the human DcR3 gene by radiation-hybrid analysis; DcR3 showed linkage to marker AFM218xe7 (T160), which maps to chromosome position 20q13. Next, we isolated from a bacterial artificial chromosome (BAC) library a human genomic clone that carries DcR3, and sequenced the ends of the clone's insert. We then determined, from the nine colon tumours that showed twofold or greater amplification of DcR3, the copy number of the DcR3-flanking sequences (reverse and forward) from the BAC, and of seven genomic markers that span chromosome 20 (Fig. 4d). The DcR3-linked reverse marker showed an average amplification of roughly threefold, slightly less than the approximately fourfold amplification of DcR3; the other markers showed little or no amplification. These data indicate that DcR3 may be at the 'epicentre' of a distal chromosome 20 region that is amplified in colon cancer, consistent with the possibility that DcR3 amplification promotes tumour survival.

Our results show that DcR3 binds specifically to FasL and inhibits FasL activity. We did not detect DcR3 binding to several other TNF-ligand-family members; however, this does not rule out the possibility that DcR3 interacts with other ligands, as do some other TNFR family members, including OPG^{2,19}.

FasL is important in regulating the immune response; however, little is known about how FasL function is controlled. One mechanism involves the molecule cFLIP, which modulates apoptosis signalling downstream of Fas²⁰. A second mechanism involves proteolytic shedding of FasL from the cell surface¹⁷. DcR3 competes with Fas for

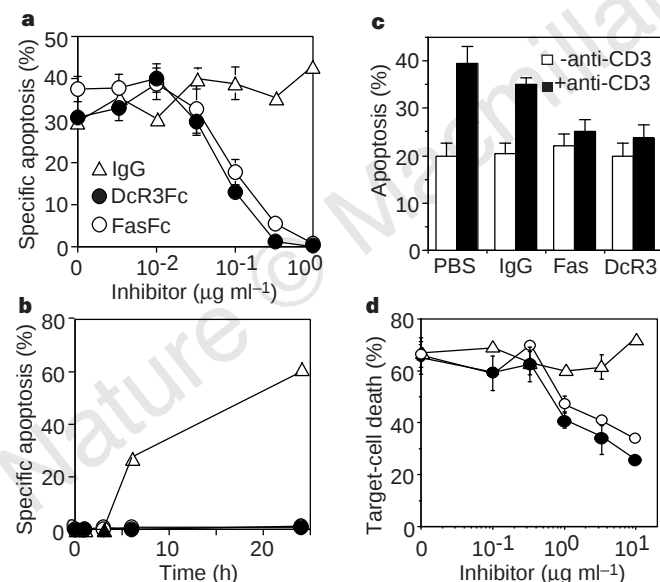


Figure 3 Inhibition of FasL activity by DcR3. **a**, Human Jurkat T leukaemia cells were incubated with Flag-tagged soluble FasL (sFasL; 5 ng ml⁻¹) oligomerized with anti-Flag antibody (0.1 μg ml⁻¹) in the presence of the proposed inhibitors DcR3-Fc, Fas-Fc or human IgG1 and assayed for apoptosis (mean ± s.e.m. of triplicates). **b**, Jurkat cells were incubated with sFasL-Flag plus anti-Flag antibody as in **a**, in presence of 1 μg ml⁻¹ DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and apoptosis was determined at the indicated time points. **c**, Peripheral blood T cells were stimulated with PHA and interleukin-2, followed by control (white bars) or anti-CD3 antibody (filled bars), together with phosphate-buffered saline (PBS), human IgG1, Fas-Fc, or DcR3-Fc (10 μg ml⁻¹). After 16 h, apoptosis of CD4⁺ cells was determined (mean ± s.e.m. of results from five donors). **d**, Peripheral blood natural killer cells were incubated with ⁵¹Cr-labelled Jurkat cells in the presence of DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and target-cell death was determined by release of ⁵¹Cr (mean ± s.d. for two donors, each in triplicate).

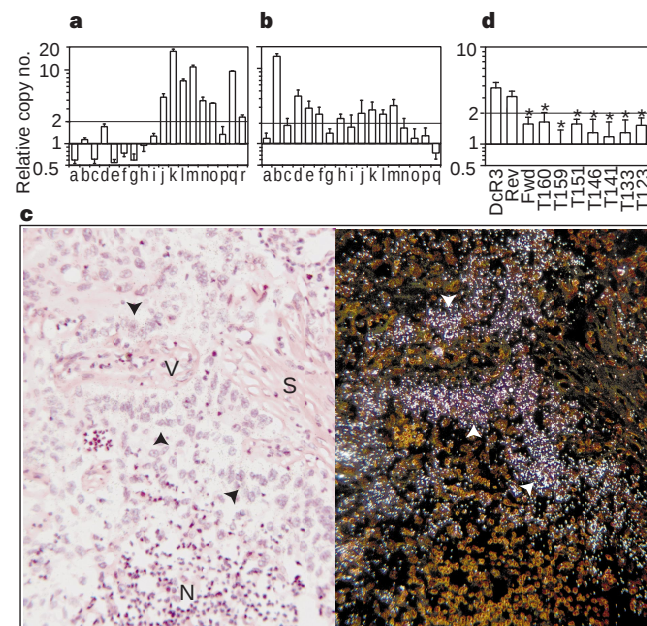


Figure 4 Genomic amplification of DcR3 in tumours. **a**, Lung cancers, comprising eight adenocarcinomas (c, d, f, g, h, j, k, r), seven squamous-cell carcinomas (a, e, m, n, o, p, q), one non-small-cell carcinoma (b), one small-cell carcinoma (i), and one bronchial adenocarcinoma (l). The data are means ± s.d. of 2 experiments done in duplicate. **b**, Colon tumours, comprising 17 adenocarcinomas. Data are means ± s.e.m. of five experiments done in duplicate. **c**, *In situ* hybridization analysis of DcR3 mRNA expression in a squamous-cell carcinoma of the lung. A representative bright-field image (left) and the corresponding dark-field image (right) show DcR3 mRNA over infiltrating malignant epithelium (arrowheads). Adjacent non-malignant stroma (S), blood vessel (V) and necrotic tumour tissue (N) are also shown. **d**, Average amplification of DcR3 compared with amplification of neighbouring genomic regions (reverse and forward, Rev and Fwd), the DcR3-linked marker T160, and other chromosome-20 markers, in the nine colon tumours showing DcR3 amplification of twofold or more (**b**). Data are from two experiments done in duplicate. Asterisk indicates $P < 0.01$ for a Student's *t*-test comparing each marker with DcR3.

FasL binding; hence, it may represent a third mechanism of extracellular regulation of FasL activity. A decoy receptor that modulates the function of the cytokine interleukin-1 has been described²¹. In addition, two decoy receptors that belong to the TNFR family, DcR1 and DcR2, regulate the FasL-related apoptosis-inducing molecule Apo2L²². Unlike DcR1 and DcR2, which are membrane-associated proteins, DcR3 is directly secreted into the extracellular space. One other secreted TNFR-family member is OPG³, which shares greater sequence homology with DcR3 (31%) than do DcR1 (17%) or DcR2 (19%); OPG functions as a third decoy for Apo2L¹⁹. Thus, DcR3 and OPG define a new subset of TNFR-family members that function as secreted decoys to modulate ligands that induce apoptosis. Pox viruses produce soluble TNFR homologues that neutralize specific TNF-family ligands, thereby modulating the antiviral immune response². Our results indicate that a similar mechanism, namely, production of a soluble decoy receptor for FasL, may contribute to immune evasion by certain tumours. □

Methods

Isolation of DcR3 cDNA. Several overlapping ESTs in GenBank (accession numbers AA025672, AA025673 and W67560) and in LifeseqTM (Incyte Pharmaceuticals; accession numbers 1339238, 1533571, 1533650, 1542861, 1789372 and 2207027) showed similarity to members of the TNFR family. We screened human cDNA libraries by PCR with primers based on the region of EST consensus; fetal lung was positive for a product of the expected size. By hybridization to a PCR-generated probe based on the ESTs, one positive clone (DNA30942) was identified. When searching for potential alternatively spliced forms of DcR3 that might encode a transmembrane protein, we isolated 50 more clones; the coding regions of these clones were identical in size to that of the initial clone (data not shown).

Fc-fusion proteins (immunoadhesins). The entire DcR3 sequence, or the ectodomain of Fas or TNFR1, was fused to the hinge and Fc region of human IgG1, expressed in insect SF9 cells or in human 293 cells, and purified as described²³.

Fluorescence-activated cell sorting (FACS) analysis. We transfected 293 cells using calcium phosphate or Effectene (Qiagen) with pRK5 vector or pRK5 encoding full-length human FasL⁴ (2 µg), together with pRK5 encoding CrmA (2 µg) to prevent cell death. After 16 h, the cells were incubated with biotinylated DcR3–Fc or TNFR1–Fc and then with phycoerythrin-conjugated streptavidin (GibcoBRL), and were assayed by FACS. The data were analysed by Kolmogorov–Smirnov statistical analysis. There was some detectable staining of vector-transfected cells by DcR3–Fc; as these cells express little FasL (data not shown), it is possible that DcR3 recognized some other factor that is expressed constitutively on 293 cells.

Immunoprecipitation. Human 293 cells were transfected as above, and metabolically labelled with [³⁵S]cysteine and [³⁵S]methionine (0.5 mCi; Amersham). After 16 h of culture in the presence of z-VAD-fmk (10 µM), the medium was immunoprecipitated with DcR3–Fc, Fas–Fc or TNFR1–Fc (5 µg), followed by protein A–Sepharose (Repligen). The precipitates were resolved by SDS–PAGE and visualized on a phosphorimager (Fuji BAS2000). Alternatively, purified, Flag-tagged soluble FasL (1 µg) (Alexis) was incubated with each Fc-fusion protein (1 µg), precipitated with protein A–Sepharose, resolved by SDS–PAGE and visualized by immunoblotting with rabbit anti-FasL antibody (Oncogene Research).

Analysis of complex formation. Flag-tagged soluble FasL (25 µg) was incubated with buffer or with DcR3–Fc (40 µg) for 1.5 h at 24 °C. The reaction was loaded onto a Superdex 200 HR 10/30 column (Pharmacia) and developed with PBS; 0.6-ml fractions were collected. The presence of DcR3–Fc–FasL complex in each fraction was analysed by placing 100 µl aliquots into microtitre wells precoated with anti-human IgG (Boehringer) to capture DcR3–Fc, followed by detection with biotinylated anti-Flag antibody Bio M2 (Kodak) and streptavidin–horseradish peroxidase (Amersham). Calibration of the column indicated an apparent relative molecular mass of the complex of 420K (data not shown), which is consistent with a stoichiometry of two DcR3–Fc homodimers to two soluble FasL homotrimers.

Equilibrium binding analysis. Microtitre wells were coated with anti-human

IgG, blocked with 2% BSA in PBS. DcR3–Fc or Fas–Fc was added, followed by serially diluted Flag-tagged soluble FasL. Bound ligand was detected with anti-Flag antibody as above. In the competition assay, Fas–Fc was immobilized as above, and the wells were blocked with excess IgG1 before addition of Flag-tagged soluble FasL plus DcR3–Fc.

T-cell AICD. CD3⁺ lymphocytes were isolated from peripheral blood of individual donors using anti-CD3 magnetic beads (Miltenyi Biotech), stimulated with phytohaemagglutinin (PHA; 2 µg ml^{−1}) for 24 h, and cultured in the presence of interleukin-2 (100 U ml^{−1}) for 5 days. The cells were plated in wells coated with anti-CD3 antibody (Pharmingen) and analysed for apoptosis 16 h later by FACS analysis of annexin-V-binding of CD4⁺ cells²⁴.

Natural killer cell activity. Natural killer cells were isolated from peripheral blood of individual donors using anti-CD56 magnetic beads (Miltenyi Biotech), and incubated for 16 h with ⁵¹Cr-loaded Jurkat cells at an effector-to-target ratio of 1:1 in the presence of DcR3–Fc, Fas–Fc or human IgG1. Target-cell death was determined by release of ⁵¹Cr in effector–target cocultures relative to release of ⁵¹Cr by detergent lysis of equal numbers of Jurkat cells.

Gene-amplification analysis. Surgical specimens were provided by J. Kern (lung tumours) and P. Quirke (colon tumours). Genomic DNA was extracted (Qiagen) and the concentration was determined using Hoechst dye 33258 intercalation fluorometry. Amplification was determined by quantitative PCR¹⁸ using a TaqMan instrument (ABI). The method was validated by comparison of PCR and Southern hybridization data for the Myc and HER-2 oncogenes (data not shown). Gene-specific primers and fluorogenic probes were designed on the basis of the sequence of DcR3 or of nearby regions identified on a BAC carrying the human DcR3 gene; alternatively, primers and probes were based on Stanford Human Genome Center marker AFM218xe7 (T160), which is linked to DcR3 (likelihood score = 5.4), SHGC-36268 (T159), the nearest available marker which maps to ~500 kilobases from T160, and five extra markers that span chromosome 20. The DcR3-specific primer sequences were 5′-CTTCTTCGCGCAGCTG-3′ and 5′-ATCACGCCGCGACCAG-3′ and the fluorogenic probe sequence was 5′-(FAM-ACACGATGCGTGCTCCAAGCAG AAp-(TAMARA), where FAM is 5′-fluorescein phosphoramidite. Relative gene-copy numbers were derived using the formula 2^(ΔCT), where ΔCT is the difference in amplification cycles required to detect DcR3 in peripheral blood lymphocyte DNA compared to test DNA.

Received 24 September; accepted 6 November 1998.

- Nagata, S. Apoptosis by death factor. *Cell* **88**, 355–365 (1997).
- Smith, C. A., Farrah, T. & Goodwin, R. G. The TNF receptor superfamily of cellular and viral proteins: activation, costimulation, and death. *Cell* **76**, 959–962 (1994).
- Simonet, W. S. *et al.* Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* **89**, 309–319 (1997).
- Suda, T., Takahashi, T., Golstein, P. & Nagata, S. Molecular cloning and expression of Fas ligand, a novel member of the TNF family. *Cell* **75**, 1169–1178 (1993).
- Pennica, D. *et al.* Human tumour necrosis factor: precursor structure, expression and homology to lymphotoxin. *Nature* **312**, 724–729 (1984).
- Pitti, R. M. *et al.* Induction of apoptosis by Apo-2 ligand, a new member of the tumor necrosis factor receptor family. *J. Biol. Chem.* **271**, 12687–12690 (1996).
- Wiley, S. R. *et al.* Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity* **3**, 673–682 (1995).
- Marsters, S. A. *et al.* Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* **8**, 525–528 (1998).
- Chicheportiche, Y. *et al.* TWEAK, a new secreted ligand in the TNF family that weakly induces apoptosis. *J. Biol. Chem.* **272**, 32401–32410 (1997).
- Wong, B. R. *et al.* TRANCE is a novel ligand of the TNFR family that activates c-Jun-N-terminal kinase in T cells. *J. Biol. Chem.* **272**, 25190–25194 (1997).
- Anderson, D. M. *et al.* A homolog of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* **390**, 175–179 (1997).
- Lacey, D. L. *et al.* Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. *Cell* **93**, 165–176 (1998).
- Dhein, J., Walczak, H., Baumler, C., Debatin, K. M. & Krammer, P. H. Autocrine T-cell suicide mediated by Apo1/(Fas/CD95). *Nature* **373**, 438–441 (1995).
- Arase, H., Arase, N. & Saito, T. Fas-mediated cytotoxicity by freshly isolated natural killer cells. *J. Exp. Med.* **181**, 1235–1238 (1995).
- Medvedev, A. E. *et al.* Regulation of Fas and Fas ligand expression in NK cells by cytokines and the involvement of Fas ligand in NK/LAK cell-mediated cytotoxicity. *Cytokine* **9**, 394–404 (1997).
- Moretta, A. Mechanisms in cell-mediated cytotoxicity. *Cell* **90**, 13–18 (1997).
- Tanaka, M., Itai, T., Adachi, M. & Nagata, S. Downregulation of Fas ligand by shedding. *Nature Med.* **4**, 31–36 (1998).
- Gelmini, S. *et al.* Quantitative PCR-based homogeneous assay with fluorogenic probes to measure c-erbB-2 oncogene amplification. *Clin. Chem.* **43**, 752–758 (1997).
- Emery, J. G. *et al.* Osteoprotegerin is a receptor for the cytotoxic ligand TRAIL. *J. Biol. Chem.* **273**, 14363–14367 (1998).
- Wallach, D. Placing death under control. *Nature* **388**, 123–125 (1997).
- Collotta, F. *et al.* Interleukin-1 type II receptor: a decoy target for IL-1 that is regulated by IL-4. *Science* **261**, 472–475 (1993).

22. Ashkenazi, A. & Dixit, V. M. Death receptors: signaling and modulation. *Science* **281**, 1305–1308 (1998).
23. Ashkenazi, A. & Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* **9**, 195–200 (1997).
24. Marsters, S. *et al.* Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* **6**, 750–752 (1996).

Acknowledgements. We thank C. Clark, D. Pennica and V. Dixit for comments, and J. Kern and P. Quirke for tumour specimens.

Correspondence and requests for materials should be addressed to A.A. (e-mail: aa@gene.com). The GenBank accession number for the DcR3 cDNA sequence is AF104419.

Crystal structure of the ATP-binding subunit of an ABC transporter

Li-Wei Hung*, Iris Xiaoyan Wang†, Kishiko Nikaido†, Pei-Qi Liu†, Giovanna Ferro-Luzzi Ames† & Sung-Hou Kim*‡

* E. O. Lawrence Berkeley National Laboratory, † Department of Molecular and Cell Biology, and ‡ Department of Chemistry, University of California at Berkeley, Berkeley, California 94720, USA

ABC transporters (also known as traffic ATPases) form a large family of proteins responsible for the translocation of a variety of compounds across membranes of both prokaryotes and eukaryotes¹. The recently completed *Escherichia coli* genome sequence revealed that the largest family of paralogous *E. coli* proteins is composed of ABC transporters². Many eukaryotic proteins of medical significance belong to this family, such as the cystic fibrosis transmembrane conductance regulator (CFTR), the P-glycoprotein (or multidrug-resistance protein) and the heterodimeric transporter associated with antigen processing (Tap1–Tap2). Here we report the crystal structure at 1.5 Å resolution of HisP, the ATP-binding subunit of the histidine permease, which is an ABC transporter from *Salmonella typhimurium*. We correlate the details of this structure with the biochemical, genetic and biophysical properties of the wild-type and several mutant HisP proteins. The structure provides a basis for understanding properties of ABC transporters and of defective CFTR proteins.

ABC transporters contain four structural domains: two nucleotide-binding domains (NBDs), which are highly conserved throughout the family, and two transmembrane domains¹. In prokaryotes these domains are often separate subunits which are assembled into a membrane-bound complex; in eukaryotes the domains are generally fused into a single polypeptide chain. The periplasmic histidine permease of *S. typhimurium* and *E. coli*^{1,3–8} is a well-characterized ABC transporter that is a good model for this superfamily. It consists of a membrane-bound complex, HisQMP₂, which comprises integral membrane subunits, HisQ and HisM, and two copies of HisP, the ATP-binding subunit. HisP, which has properties intermediate between those of integral and peripheral membrane proteins⁹, is accessible from both sides of the membrane, presumably by its interaction with HisQ and HisM⁶. The two HisP subunits form a dimer, as shown by their cooperativity in ATP hydrolysis⁵, the requirement for both subunits to be present for activity⁸, and the formation of a HisP dimer upon chemical cross-linking. Soluble HisP also forms a dimer³. HisP has been purified and characterized in an active soluble form³ which can be reconstituted into a fully active membrane-bound complex⁸.

The overall shape of the crystal structure of the HisP monomer is that of an 'L' with two thick arms (arm I and arm II); the ATP-binding pocket is near the end of arm I (Fig. 1). A six-stranded β -sheet (β 3 and β 8– β 12) spans both arms of the L, with a domain of a α - plus β -type structure (β 1, β 2, β 4– β 7, α 1 and α 2) on one side (within arm I) and a domain of mostly α -helices (α 3– α 9) on the

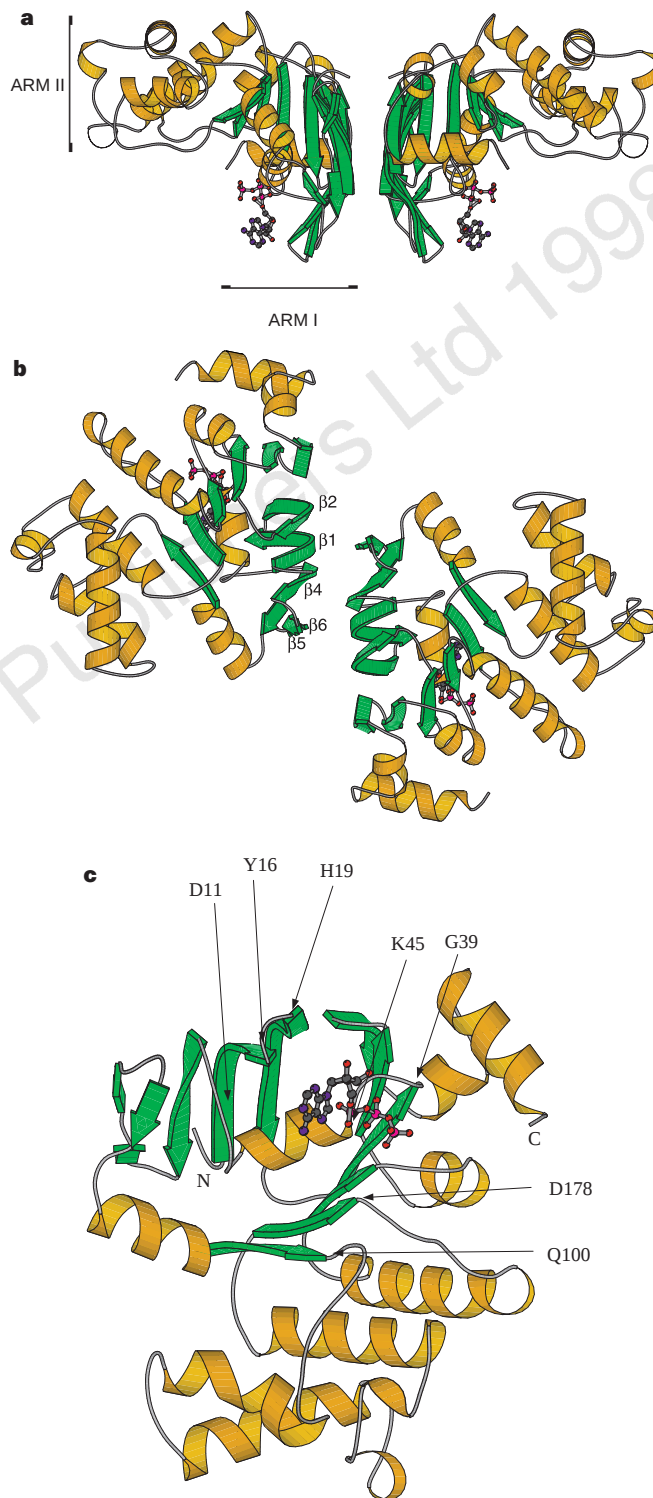


Figure 1 Crystal structure of HisP. **a**, View of the dimer along an axis perpendicular to its two-fold axis. The top and bottom of the dimer are suggested to face towards the periplasmic and cytoplasmic sides, respectively (see text). The thickness of arm II is about 25 Å, comparable to that of membrane. α -Helices are shown in orange and β -sheets in green. **b**, View along the two-fold axis of the HisP dimer, showing the relative displacement of the monomers not apparent in **a**. The β -strands at the dimer interface are labelled. **c**, View of one monomer from the bottom of arm I, as shown in **a**, towards arm II, showing the ATP-binding pocket. **a–c**, The protein and the bound ATP are in 'ribbon' and 'ball-and-stick' representations, respectively. Key residues discussed in the text are indicated in **c**. These figures were prepared with MOLSCRIPT²⁹. N, amino terminus; C, C terminus.

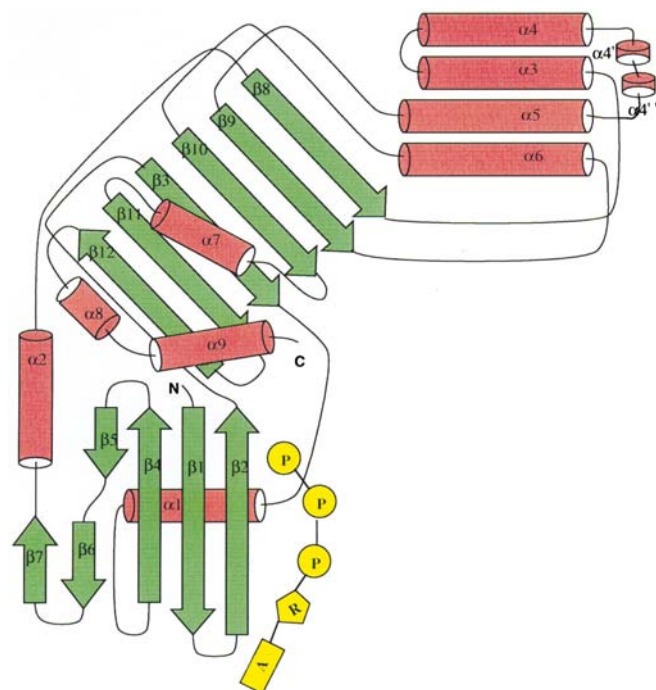


Figure 2 Diagram of the secondary structure of HisP. Cylinders and arrows represent α -helices and β -strands, respectively. ATP is shown in yellow (A, adenine; R, ribose; P, phosphate group).

other side (within arm II) of the sheet (Fig. 2). The overall fold of the structure is different from that of any known protein. However, when we searched the Protein Databank we found limited similarities of some parts of the structure with the structures of RecA¹⁰ and of the α - and β -subunits of bovine F_1 -ATPase¹¹.

The ATP-binding pocket contains the 'phosphate-binding loop' or 'P-loop' that is found in the structures of nucleotide-binding proteins such as Ras and adenylate kinase¹². This corresponds to the A motif of Walker's A/B motif¹³, and includes residues Gly 39 to Lys 45 which link $\beta 3$ to $\alpha 1$. The P-loop wraps over the β -phosphate of ATP, allowing the formation of extensive hydrogen bonding between the main-chain nitrogens of Ser 41 to Lys 45 and the β -phosphate, and interacts directly with the γ -phosphate of ATP through a hydrogen bond with Ser 41. The main-chain nitrogen of Gly 39 interacts with the side-chain amino group of Lys 45. Tyr 16 provides an additional stabilizing force towards ATP by way of an extensive hydrophobic interaction that buries 172 Å² of the surface area between its aromatic ring and the adenine base of ATP. All of the residues forming hydrogen bonds or stacking with ATP are shown in Fig. 3. Of these, Asp 178, located at the end of $\beta 9$, is particularly interesting: it interacts with the γ -phosphate of ATP through a water molecule (water 407) that occupies a position similar to that of the magnesium ion in the crystal structures of Ras-GMPPNP-Mg¹⁴ and of the $F_1\alpha_3\beta_3$ -AMPPNP-Mg complex¹¹. As the HisP structure does not contain magnesium (because the presence of magnesium induces ATP hydrolysis which destabilizes the solubility of HisP³), it is likely that water 407 occupies the position that is usually occupied by the divalent cation required for ATP hydrolysis. Asp 178, which is highly conserved in ABC transporters, is probably the residue that interacts with the divalent cation during ATP hydrolysis, by analogy with the process of GTP hydrolysis in GTPases^{15,16}. Glu 179 and Gln 100, which are also highly conserved, form hydrogen bonds with water 437, which interacts with the γ -phosphate through a hydrogen bond that is nearly parallel to the $P\gamma-O\beta$ bond, the bond that breaks upon hydrolysis. Water 437 is the most likely candidate for the 'attacking' water during ATP hydrolysis. If this were the case, either Gln 100 or Glu 179 could

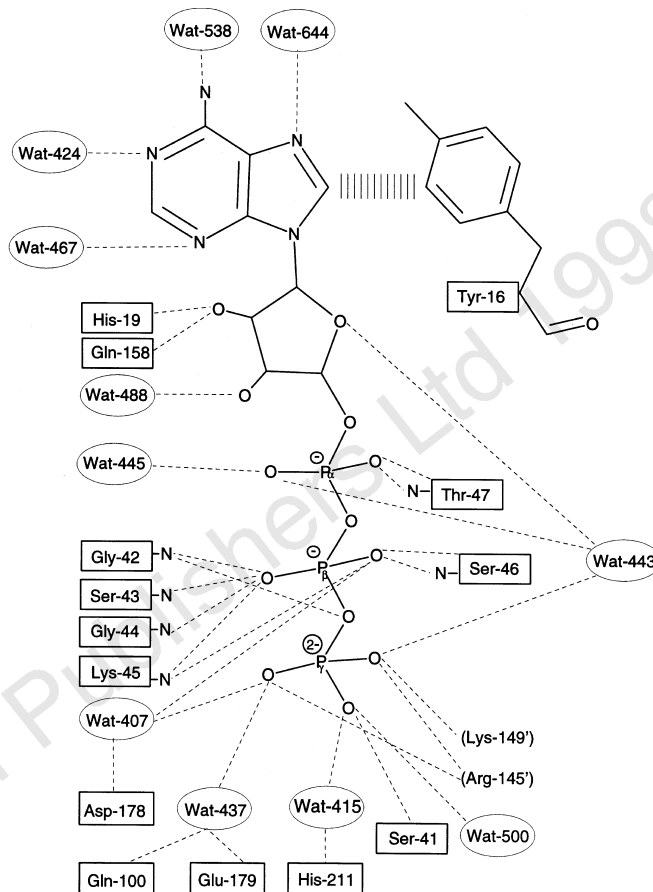


Figure 3 Atomic details of the interactions of HisP (residues are boxed) with ATP (centre). Only hydrogen bonds and aromatic stacking are shown, by dashed lines and vertical lines, respectively. Interactions of ATP with protein side chains are indicated by lines connecting to the respective residue boxes; interactions of ATP with protein main chain atoms are shown as lines to the atoms themselves. The residues in parentheses (Arg 145' and Lys 149') belong to a different, symmetry-related HisP dimer.

be the activating residue.

In agreement with the idea that HisP forms a dimer *in vivo*, a dimer consisting of two monomers related by a two-fold axis is also found in the crystal structure of HisP. Figure 1a, b shows two views of the two monomers contacting each other through hydrophobic interactions at the antiparallel β -sheet on the outer side of arm I through strands $\beta 1$, $\beta 4$ and $\beta 5$. Two buried negative charges, contributed by Asp 11 of each of the monomers, are located at the dimer interface and stabilized by intermolecular hydrogen bonds to water 587 and to the backbone of the two-fold-related HisP molecule (Leu 72'). This situation is energetically unfavourable; thus, either Asp 11 must be protonated or one of the electron densities that we interpreted to be water molecules may be a metal ion. The dimer is 60 Å thick, 40 Å tall, and 90 Å wide.

The crystal structure correlates well with many genetic and biochemical data obtained *in vivo*¹⁷. Proteins with the mutations G39D, G39R, G39K, G44S, K45P, K45V or K45N are defective in nucleotide binding, ATP hydrolysis and ligand translocation. The affected residues are located in positions in the crystal structure that appear to be important for ATP hydrolysis and/or binding; that is, these residues are located in the P-loop of the ATP-binding site or are involved in hydrogen bonding to ATP (Fig. 1c, Fig. 3). As would be expected from the crystal structure, mutation of Asp 178 (D178R and D178K) results in defective nucleotide binding and transport. Mutation in Glu 179 (E179D), which also interacts with the

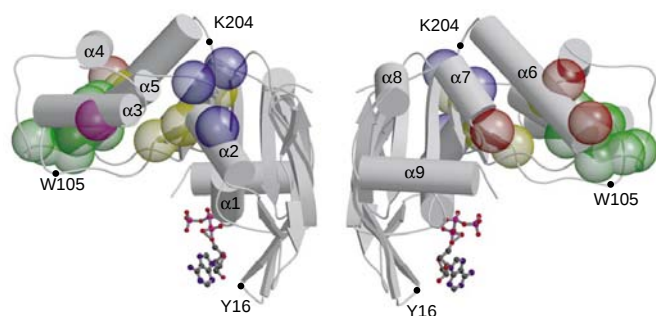


Figure 4 Locations of signal-independent HisP mutations in the crystal structure. Mutated residues of groups 1, 2 and 3 (see text) appear as blue, dark red and gold spheres, respectively. The purple (left) indicates residue 116 in HisP; this residue corresponds to the most commonly mutated CFTR residue, F508, according to sequence alignment (Fig. 5). Green spheres indicate 'motif C' (see text). The orientation of the protein is the same as in Fig. 1a. Bound ATP is shown as a ball-and-stick representation. α -Helices are numbered. This figure was prepared with raster3D (ref. 30). The positions of residues Y16, W105, K204 discussed in the text are labelled.

γ -phosphate of ATP through water 437, also eliminates transport and ATP hydrolysis, although nucleotide binding is normal. Tyr 16 is located in the crystal structure in a short stretch between β 1 and β 2; mutant Y16S is unable to bind nucleotide, which is consistent with the possibility that β 1 and β 2 undergo a conformational change upon ATP binding and/or ADP departure after hydrolysis.

Covalent crosslinking of HisP with 8-azido-ATP results in derivatization of His 19 and Ser41 (ref. 18), which in the crystal structure line the inner surface of the ATP-binding pocket. Substitution of His 19, which forms a hydrogen bond with the O2' of the ATP ribose, with glutamine has little effect on HisP's ability to bind or hydrolyse ATP or on transport proficiency¹⁷. This is compatible with the idea that the substituting glutamine side chain supplies the hydrogen bond and that this hydrogen bond is not involved in ATP hydrolysis.

A conserved glutamine-rich sequence (residues 154–162 in HisP; bacterial consensus sequence LSGGQQQRV), called the 'linker peptide'¹⁹, 'signature motif' or 'motif C'²⁰, is located in arm II in the crystal structure, with part of it (residues Gly 156 to Val 162) forming the first half of α 5 and with residues 154–158 only being partially exposed (Fig. 4). Several mutations of this motif (S155F, G157D and R161S) inactivate the permease activity, indicating that the motif performs an essential function. It is possible that the motif is essential for the integrity of the folded HisP molecule.

The conformation of HisP in solution³ is similar to that in the crystal, according to several criteria. The CD spectrum of soluble HisP indicates a secondary-structure composition of 40% α -helices and 23% β -strands, which corresponds with the values of 38% and 23%, respectively, calculated from the crystal structure. The tryptophan-fluorescence spectrum of the single tryptophan in HisP (Trp 105) has a maximum at 348 nm, indicating that Trp 105 is in a relatively polar environment. This is in agreement with the location of Trp 105 in a loop between β 8 and α 3 and with its 80% exposure. Consistent with the fact that tryptophan fluorescence is not affected either by ATP or by HisP's dimerization status, Trp 105 appears to be distant from possible conformational changes that occur either in the nucleotide-binding pocket or as a consequence of the dimerization process. The tryptophan-fluorescence maximum shifts to 337 nm upon exposure to phospholipids^{3,8}, indicating either that Trp 105 interacts directly with the phospholipids or that the loop containing Trp 105 undergoes a conformational change in the

presence of phospholipids (and thus, possibly, upon interaction with HisQ/M).

All available biochemical data about HisP in the HisQMP₂ complex are compatible with the characteristics of HisP in the crystal structure. With regard to arm I, first, soluble HisP hydrolyses ATP, indicating that the conformation of the features associated with ATP hydrolysis in arm I may be the same in the crystal and in the HisQMP₂ complex. Second, the ATP-binding pocket in the crystal structure has features in common with other nucleotide-binding proteins^{10,11}. Thus, this portion of the structure is likely to be the same whether HisP is in solution or within HisQMP₂. Third, HisP in HisQMP₂ is accessible by antibodies from the cytoplasmic side. As ATP is located in the cytoplasm, the ATP-binding pocket in arm I of HisP should be facing towards the cytoplasm as shown in the representation of the crystal structure in Fig. 1a. Fourth, HisP in HisQMP₂ is also accessible from the periplasmic side, as judged from patterns of proteolytic digestion and modification with a membrane-impermeant biotinylating reagent⁶, which modifies mainly Lys 204 (G.F.-L.A. *et al.*, unpublished data). In the crystal structure, Lys 204 is exposed and located at the opposite end of arm I from the ATP-binding site (Fig. 4). The distance between Lys 204 and the C α atom of Tyr 16 is about 33 Å, which is consistent with the accessibility of HisP at both sides of the membrane. Together these results indicate that arm I in the crystal structure of HisP may largely represent arm I of HisP in HisQMP₂.

With regard to arm II, because arms I and II share a six-stranded β -sheet (Fig. 1c, Fig. 2), and as the conformation of arm I is likely to be similar in the crystal structure of HisP and in the HisQMP₂ complex, it is probable that the conformation of most of arm II is also similar in the crystal structure and in the complex. Furthermore, HisP is embedded in the membrane by interaction with the hydrophobic subunits, HisQ and HisM⁶, and strong genetic evidence indicates that arm II is closely involved in this interaction *in vivo* (see below). This idea is compatible with the thickness of arm II being about 25 Å (Fig. 1a). In summary, the biochemical, biophysical and genetic properties of wild-type HisP and of numerous HisP mutant proteins, together with the fact that HisP is a dimer both in the crystal and in HisQMP₂, support the ideas that the crystal structure corresponds to the structure of HisP within the *in vivo* HisQMP₂ complex, that HisP can be accessed from both sides of the membrane, and that arm II is likely to be embedded in the membrane, surrounded by HisQ and HisM.

Many mutations have been characterized that release HisP from the regulatory control of HisQ/HisM and result in constitutive ATPase activity (signal-independent mutants)⁷. All of the residues involved are located in arm II (Fig. 4) and can be divided into three groups: group 1, L90S, P172T/L and T205A/M are located on the surface on one side of arm II; group 2, E187K, E191K and E212K/G are on the surface on the other side; and group 3, V98A, M206I, V192M and T96(+T) are buried. These mutant HisQMP₂ complexes generally have a looser structure than the wild type, with HisP being easily released from the complex (P.-Q.L. & G.F.-L.A., manuscript in preparation). Therefore, these mutations disrupt the interaction between HisP and HisQ/HisM and/or the tertiary conformation of arm II.

ATP hydrolysis by HisP is induced by a signal from liganded HisJ at the periplasmic surface⁷. A possible scenario is that, upon interaction with liganded HisJ, HisQ and/or HisM undergo conformational changes that cause conformational changes in HisP's arm II, which is in direct contact with HisQ/M; because arm II contains several residues (Gln 100, Asp 178 and Glu 179) that are important for ATP hydrolysis, it is possible that such changes result directly in activation of ATP hydrolysis in arm I. Alternatively, as arm II shares with arm I a six-stranded β -sheet in which Asp 178 is located, ATP hydrolysis may be activated by a conformational change within HisP, such as a change in the relative orientation of arms I and II.

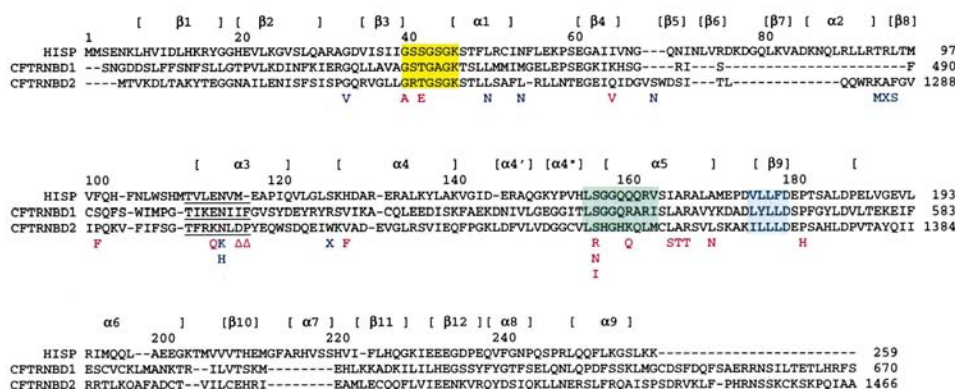


Figure 5 Sequence alignment of HisP and the NBDs of human CFTR. The top rows show secondary-structure assignments and the second rows show residue numbers of HisP. The colour shadings indicate the positions of motifs A (yellow), B (blue) and C (green). The bottom rows show common amino-acid substitutions in

NBD1 (red) and NBD2 (blue) of CFTR. The underlined residues indicate a motif that is well conserved in ABC transporters²⁰; this motif includes F508 in CFTR. Dashes represent gaps introduced to improve the sequence alignment.

Cystic fibrosis, a lethal disease that affects 1 in 2,000 people in the USA and Canada and 1 in 25 Caucasians, is caused by mutations in the CFTR gene. CFTR is a single-chain peptide of 1,480 residues and consists of five domains, two of which (NBD1 and NBD2) are nucleotide-binding domains homologous to HisP. Figure 5 shows a sequence alignment of HisP and of the two NBDs of CFTR. Using the HisP crystal structure as a model for the NBDs of CFTR, we can propose a structural basis for the defect of various CFTR mutants. The most common mutation among all cystic-fibrosis patients, $\Delta F508$ (about 90% of cystic-fibrosis patients have been reported to have at least one $\Delta F508$ CFTR allele), is located in a conserved sequence stretch and corresponds to an exposed area of helix $\alpha 3$ in HisP (Fig. 4). This region is in arm II, which we propose to be involved in interactions with the membrane-spanning domains (HisQ/HisM) and in which mutations that result in more loosely assembled complexes and constitutive ATPase activity are located. The defect in $\Delta F508$ resides in its inability to move from the endoplasmic reticulum to the plasma membrane²¹, rather than in its channel activity. By analogy with the

defects in the mutants shown in Fig. 4, it is likely that the $\Delta F508$ mutant also has defects in the interaction between NBD1 and the transmembrane domains (TMDs), which would lead to incorrect folding and transit. Many other CFTR mutations are found in regions corresponding to arm II of HisP, with most of the mutated residues being on the surface and some being buried. Mutations in arm II are likely to disrupt the interaction between NBDs and TMDs. Another group of CFTR mutations are located in or near the ATP-binding pocket and presumably alter ATP binding or hydrolysis. A third group of mutations are located on or near the proposed dimer interface of HisP ($\beta 2$ and $\beta 4$; Fig. 1b) and are likely to disrupt the formation of the NBD dimer.

In conclusion, the crystal structure of HisP is consistent with a large body of biochemical, biophysical and genetic data obtained using isolated HisP and HisP in HisQMP₂, suggesting that the structure represents a functionally relevant form of HisP *in vivo*. Furthermore, the ability to correlate the properties of CFTR mutants with the crystal structure of HisP indicates that HisP is a good model for the NBDs of ABC transporters in general.

Note added in proof: While this manuscript was in preparation, an abstract appeared in the American Crystallographic Association 1998 Annual Meeting Abstract reporting the structure of N-terminal domain RbsA, at 2.5 Å resolution.

Table 1 Data collection and refinement statistics

MAD data statistics (20.0–1.7 Å)*				
Wavelength (Å)	1	0.9800	0.9798	0.9686
1	0.032 (0.034)	0.069	0.053	0.026
0.9800		0.035 (0.037)	0.027	0.041
0.9798			0.043 (0.039)	0.052
0.9686				0.055 (0.051)
Non-hydrogen atoms in an asymmetric unit			2,363	
d-spacings (Å)			20.0–1.5	
		$F > 0\sigma$		$F > 2\sigma$
Unique reflections (free)		53,885 (5458)		49,835 (5039)
Completeness		93.3%		86.3%
R-value† (R _{free})		0.195 (0.220)		0.185 (0.211)
r.m.s.d. bonds length (Å)		0.018		
r.m.s.d. bond angle (°)		2.3		
Average B-value (Å ²)		18.2		
r.m.s.d. B-values (Å ²)		4.5		

* Values represent the absolute value of the Bijvoet difference at one wavelength (diagonal elements) or the dispersive differences between two wavelengths (off-diagonal element). The values in parentheses are the ratios for centric Bijvoet reflections.

† This model contained residues 5–262 of HisP–HisP, one ATP, three possible anions (modelled as Cl[−], and 322 solvent molecules. The coordinate has been validated with the program PROCHECK²⁸ and all geometric parameters are within acceptable or better regions.

‡ R-factor = $\sum_{hkl} ||F_o(hkl)| - |F_c(hkl)|| / \sum_{hkl} |F_o(hkl)|$

$R_{free} = \sum_{hkl} ||F_o(hkl, test)| - |F_c(hkl, test)|| / \sum_{hkl} |F_o(hkl, test)|$

F_o is observed structure factor; F_c is calculated structure factor based on the model; and test is reflections omitted during refinement.

r.m.s.d., root mean square deviation.

Methods

His₆-tagged His was purified as described³. Crystallization conditions were established using the sparse matrix method. Crystals appeared from a solution containing 15 mg ml^{−1} protein, 20% glycerol, 2 mM EDTA, 50 mM imidazole, 0.1 M HEPES buffer, pH 7.0, and 10 mM ATP by mixing this solution (1:1) with crystallization buffer (20% polyethyleneglycol (PEG) 6K, 1.2 M LiCl, 0.1 M HEPES buffer, pH 7.0, and 10 mM ATP) and equilibrating it against crystallization buffer. Crystals (~0.8 mm × 0.4 mm × 0.2 mm) diffracted beyond 2.1 Å on a CuKα rotating anode generator and better than 1.5 Å with 1.0 Å synchrotron radiation source. HisP crystallizes in space group $P4_32_12$ with cell parameters 67 × 67 × 149 Å³. Selenomethionine-substituted HisP was purified and crystallized as for the wild-type protein, except that the best crystals appeared in the crystallization buffer containing 1.0 M LiCl instead of 1.2 M LiCl; crystals were of similar size and had the same diffraction quality as native crystals and electrospray mass-spectrographic data of selenomethionine-substituted HisP indicate that all methionines had been selenomethionine-substituted.

Both multiple isomorphous replacement (MIR) (using 2 Hg, 1 Ir and 1 Se derivatives) and multiple anomalous diffraction (MAD) methods gave high-quality maps. The final model was built into the electron-density map phased by the MAD method because the resolution of the MAD data set was much higher. For MAD data collection, one crystal of selenomethionine HisP was transferred into a cryosolvent containing 20% glycerol, 20% PEG 6K, 0.2M LiCl, 0.1M HEPES buffer, pH 7.0, and 10 mM ATP for 2 s, flash-frozen in liquid nitrogen, and transferred into a cryostream set at 100 K. A four-wavelength

MAD data set was collected at the Advanced Light Source in 2 h. Diffraction data were processed with DENZO and SCALEPACK²². The program SOLVE²³ was used to perform the search for heavy atoms and for phasing. Solvent flattening and histogram matching with program DM²⁴ were used to improve the map quality. The initial model was built with program O²⁵ and yielded a continuous polypeptide chain between residues 5 and 262. The model was refined with package CNS²⁶ against the maximum-likelihood target²⁷ with bulk-solvent correction. Crystallographic data are summarized in Table 1.

Received 4 September; accepted 27 November 1998.

- Doige, C. A. & Ames, G. F.-L. ATP-dependent transport systems in bacteria and humans: relevance to cystic fibrosis and multidrug resistance. *Annu. Rev. Microbiol.* **47**, 291–319 (1993).
- Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474 (1997).
- Nikaido, K., Liu, P.-Q. & Ames, G. F.-L. Purification and characterization of HisP, the ATP-binding subunit of a traffic ATPase (ABC transporter), the histidine permease of *Salmonella typhimurium*. Solubilization, dimerization, and ATPase activity. *J. Biol. Chem.* **272**, 27745–27752 (1997).
- Liu, C. E. & Ames, G. F.-L. Characterization of transport through the periplasmic histidine permease using proteoliposomes reconstituted by dialysis. *J. Biol. Chem.* **272**, 859–866 (1997).
- Liu, C. E., Liu, P.-Q. & Ames, G. F.-L. Characterization of the adenosine triphosphatase activity of the periplasmic histidine permease, a traffic ATPase (ABC transporter). *J. Biol. Chem.* **272**, 21883–21891 (1997).
- Baichwal, V., Liu, D. & Ames, G. F.-L. The ATP-binding component of a prokaryotic traffic ATPase is exposed to the periplasmic (external) surface. *Proc. Natl Acad. Sci. USA* **90**, 620–624 (1993).
- Petronilli, V. & Ames, G. F.-L. Binding protein-independent histidine permease mutants: uncoupling of ATP hydrolysis from transmembrane signaling. *J. Biol. Chem.* **266**, 16293–16296 (1991).
- Liu, P.-Q. & Ames, G. F.-L. *In vitro* disassembly and reassembly of an ABC transporter, the histidine permease. *Proc. Natl Acad. Sci. USA* **95**, 3495–3500 (1998).
- Kerppola, R. E., Shyamala, V., Klebba, P. & Ames, G. F.-L. The membrane-bound proteins of periplasmic permeases form a complex: identification of the histidine permease HisQMP complex. *J. Biol. Chem.* **266**, 9857–9865 (1991).
- Story, R. M. & Steitz, T. A. Structure of the recA protein-ADP complex. *Nature* **355**, 318–324 (1992).
- Abrahams, J. P., Leslie, A. G., Lutter, R. & Walker, J. E. Structure at 2.8 Å resolution of F₁-ATPase from bovine heart mitochondria. *Nature* **370**, 621–628 (1994).
- Saraste, M., Sibbald, P. R. & Wittinghofer, A. The P-loop—a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430–434 (1990).
- Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the α and β -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945–951 (1982).
- Pai, E. F. *et al.* Structure of the guanine-nucleotide-binding domain of the Ha-ras oncogene product p21 in the triphosphate conformation. *Nature* **341**, 209–214 (1989).
- Kim, S.-H., Prive, G. G. & Milburn, M. V. in *Handbook of Experimental Pharmacology. GTPases in Biology* (eds Dickey, B. F. & Birnbaumer, L.) 177–194 (Springer, Berlin, 1993).
- Pai, E. F. *et al.* Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 Å resolution: implications for the mechanism of GTP hydrolysis. *EMBO J.* **9**, 2351–2359 (1990).
- Shyamala, V., Baichwal, V., Beall, E. & Ames, G. F.-L. Structure-function analysis of the histidine permease and comparison with cystic fibrosis mutations. *J. Biol. Chem.* **266**, 18714–18719 (1991).
- Mimura, C. S., Admon, A., Hurt, K. A. & Ames, G. F.-L. The nucleotide-binding site of HisP, a membrane protein of the histidine permease. Identification of amino acid residues photoaffinity-labeled by 8-azido ATP. *J. Biol. Chem.* **265**, 19535–19542 (1990).
- Ames, G. F.-L., Mimura, C., Holbrook, S. & Shyamala, V. Traffic ATPases: a superfamily of transport proteins operating from *Escherichia coli* to humans. *Adv. Enzymol.* **65**, 1–47 (1992).
- Bianchet, M. A., Ko, Y. H., Amzel, L. M. & Pedersen, P. L. Modeling of nucleotide binding domains of ABC transporter proteins based on a F₁-ATPase/recA topology: structural model of the nucleotide binding domains of the cystic fibrosis transmembrane conductance regulator (CFTR). *J. Bioenerg. Biomemb.* **29**, 503–524 (1997).
- Cheng, S. H. *et al.* Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell* **63**, 827–834 (1990).
- Otwinski, Z. in *Data Collection and Processing* (eds Saway, L., Isaacs, N. & Bailey, S.) 56–62 (SERC Daresbury Laboratory, Warrington, UK, 1993).
- Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. Generalized method of determining heavy-atom positions using the difference Patterson function. *Acta Crystallogr. A* **43**, 1–5 (1987).
- Cowan, K. D. Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. *Acta Crystallogr. D* **49**, 148–157 (1993).
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. Cross-validated maximum likelihood enhance crystallographic simulated annealing refinement. *Proc. Natl Acad. Sci. USA* **94**, 5018–5023 (1997).
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. Procheck—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
- Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).

Acknowledgements. We thank J. Jancarik for crystallization of native HisP crystals, and D. S. King for carrying out mass-spectroscopy analysis on the native and Se-Met HisP proteins. The MAD data for this study were collected at the Lawrence Berkeley National Laboratory in the Macromolecular Crystallography Facility at beamline 5.0.2 in the Advanced Light Source. This facility is principally funded by the Office of Biological and Environmental Research of the US Department of Energy Offices with contributions from Lawrence Berkeley National Laboratory, Amgen, Roche Biosciences, the University of California, Berkeley, and Lawrence Livermore National Laboratory. This work has been supported by grants from the Health Effects and Life Sciences Research Division, Office of Biosciences and Environmental Research, Office of Energy Research of Department of Energy (to S.-H.K.) and from the NIH (to G.F.-L.A.).

Correspondence and requests for materials should be addressed to S.-H.K. (e-mail: shkim@lbl.gov). The atomic coordinate has been deposited at the Brookhaven Protein Databank under code 1b0u.

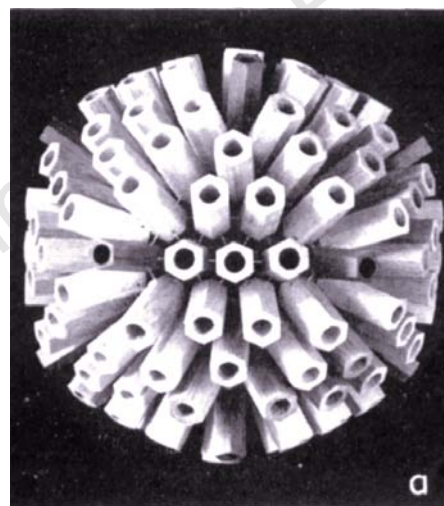
erratum

Visible viruses

Martin Kemp

Nature **396**, 123 (1998)

Readers of some copies of the 12 November issue of *Nature* will have found nothing but a black rectangle in the 'Science and Image' section where Fig. 2 was meant to be. What they should have seen is "a neat wooden model" of a herpes virus, as shown here. □



correction

Crystal structure of a bacterial signal peptidase in complex with a β -lactam inhibitor

Mark Paetzel, Ross E. Dalbey & Natalie C. J. Strynadka

Nature **396**, 186–190 (1998)

We omitted to give details of the coordinates of the bacterial signal peptidase: these have been deposited with the Protein Data Bank (code 1B12). Because our Letter was submitted before the 1 October 1998 deadline (see editorial in *Nature* **394**, 105; 1998), these coordinates will be held at the PDB for one year before release. □